

OM protein - protein search, using sw model

Run on: January 7, 2005, 11:40:59 ; Search time 67.1545 Seconds  
 (without alignments)  
 1864.305 Million cell updates/sec

Title: US-10-726-721A-7  
 Perfect score: 1921  
 Sequence: 1 VLGACNAVNYAADNQIYIAG.....DFSESRRNRFSTPEQAAKNR 349

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : A\_Geneseq\_23Sep04:\*  
 1: geneseqp1980s:\*  
 2: geneseqp1990s:\*  
 3: geneseqp2000s:\*  
 4: geneseqp2001s:\*  
 5: geneseqp2002s:\*  
 6: geneseqp2003as:\*  
 7: geneseqp2003bs:\*  
 8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query Match	Length			
1	1921	100.0	349	4	AAG67775	Aag67775 Amino aci
2	1921	100.0	589	4	AAU27959	Aau27959 Human con
3	1921	100.0	589	4	AAU27787	Aau27787 Human ful
4	1909	99.4	558	6	ABO52974	Abo52974 Human spl
5	1909	99.4	589	5	ABB97144	Abb97144 Human tum
6	1897	98.8	567	4	ABG15420	Abg15420 Novel hum
7	1812	94.3	624	4	AAU33004	Aau33004 Novel hum
8	1286.5	67.0	379	4	AAU33002	Aau33002 Novel hum
9	1286.5	67.0	404	4	ABG15417	Abg15417 Novel hum

10	976.5	50.8	437	5	ABP43680	Abp43680	Human	RNA
11	976.5	50.8	537	3	AAy70236	Aay70236	Human	RNA
12	976.5	50.8	537	3	AAB41893	Aab41893	Human	ORF
13	976.5	50.8	542	7	ADI63130	Adi63130	Human	apo
14	976.5	50.8	565	4	ADM20004	Adm20004	Protein	e
15	780	40.6	301	3	AAB43909	Aab43909	Human	can
16	671	34.9	549	4	ABG00852	Abg00852	Novel	hum
17	671	34.9	686	4	ABG00854	Abg00854	Novel	hum
18	655	34.1	475	4	ABB64307	Abb64307	Drosophil	
19	653	34.0	152	4	AAO04402	Aao04402	Human	pol
20	525	27.3	290	4	ADM19741	Adm19741	Protein	e
21	439	22.9	140	4	AAU33003	Aau33003	Novel	hum
22	426	22.2	168	7	ADM06009	Adm06009	Human	pro
23	391	20.4	373	3	AAG48637	Aag48637	Arabidops	
24	391	20.4	384	3	AAG48636	Aag48636	Arabidops	
25	391	20.4	420	3	AAG48635	Aag48635	Arabidops	
26	382	19.9	383	3	AAG25383	Aag25383	Arabidops	
27	382	19.9	394	3	AAG25382	Aag25382	Arabidops	
28	382	19.9	430	3	AAG25381	Aag25381	Arabidops	
29	380.5	19.8	392	3	AAG48625	Aag48625	Arabidops	
30	380.5	19.8	403	3	AAG48624	Aag48624	Arabidops	
31	380.5	19.8	439	3	AAG48623	Aag48623	Arabidops	
32	353	18.4	557	7	ABR84745	Abr84745	Human	pol
33	353	18.4	557	7	ADF69120	Adf69120	Human	MP5
34	349.5	18.2	550	5	AAU80386	Aau80386	Human	lun
35	349.5	18.2	550	7	ADB95124	Adb95124	Human	lun
36	343.5	17.9	532	7	ADB79923	Adb79923	Rat	PTB-1
37	343	17.9	531	6	ABO52966	Abo52966	Human	spl
38	343	17.9	531	8	ADN04554	Adn04554	Antipsori	
39	342.5	17.8	531	7	ADF69118	Adf69118	Human	MP5
40	342.5	17.8	532	4	AAM40048	Aam40048	Human	pol
41	341	17.8	568	4	ABB58853	Abb58853	Drosophil	
42	329.5	17.2	521	3	AAB42270	Aab42270	Human	ORF
43	329.5	17.2	521	7	ADF69119	Adf69119	Human	MP5
44	324	16.9	322	7	ADJ70115	Adj70115	Human	hea
45	324	16.9	345	4	AAB95136	Aab95136	Human	pro

# ALIGNMENTS

## RESULT 1

AAG67775

ID AAG67775 standard; protein; 349 AA.

XX

AC AAG67775;

XX

DT 21-JAN-2002 (first entry)

XX

DE Amino acid sequence of a human hnRNPL protein.

XX

KW Human; phosphotyrosine binding domain 1; PTB1 domain; FE65; beta-amyloid;

KW Alzheimer's disease; FEBP1; FE65 binding PTB1 domain protein; hnRNPL;

KW neurodegenerative disease.

XX

OS Homo sapiens.

XX

PN WO200159104-A1.  
 XX  
 PD 16-AUG-2001.  
 XX  
 PF 07-FEB-2001; 2001WO-FR000361.  
 XX  
 PR 10-FEB-2000; 2000FR-00001628.  
 PR 18-APR-2000; 2000US-0198500P.  
 XX  
 PA (AVET ) AVENTIS PHARMA SA.  
 XX  
 PI Maury I, Mercken L, Fournier A;  
 XX  
 DR WPI; 2001-589717/66.  
 DR N-PSDB; AAH78614.  
 XX  
 PT Compound capable of modulating interaction between the PTB1 domain of  
 PT FE65 protein and hnRNPL and/or FEBP1 protein, useful to treat  
 PT neurological disorders including Alzheimer's disease.  
 XX  
 PS Claim 10; Page 39-40; 51pp; French.  
 XX  
 CC The present sequence represents a human hnRNPL (undefined) protein. The  
 CC protein is a partner of the human FE65 protein. FE65 is implicated in the  
 CC production of beta-amyloid. Partners of the FE65 protein thus represent  
 CC novel targets for the treatment of Alzheimer's disease. Such partners  
 CC include FEBP1 (FE65 binding PTB1 domain protein) and hnRNPL (undefined).  
 CC Compounds which are capable of at least partially modulating interactions  
 CC between hnRNPL and/or FEBP1 proteins or their homologues and the  
 CC phosphotyrosine binding domain 1 (PTB1) domain of FE65 are used to treat  
 CC neurodegenerative diseases. In particular, they are used for treating  
 CC Alzheimer's disease  
 XX  
 SQ Sequence 349 AA;

Query Match 100.0%; Score 1921; DB 4; Length 349;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-169;  
 Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDDSRSVNSVLLFTILNPIYSI	60
Db	1	VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDDSRSVNSVLLFTILNPIYSI	60
Qy	61	TTDLVLYTICNPCGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE	120
Db	61	TTDLVLYTICNPCGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE	120
Qy	121	YAKPTRLNVFKNDQDTWDYTNPNLSGQDPSNPNKRQRQPPLLGDHPAEYGGPHGGYHS	180
Db	121	YAKPTRLNVFKNDQDTWDYTNPNLSGQDPSNPNKRQRQPPLLGDHPAEYGGPHGGYHS	180
Qy	181	HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV	240
Db	181	HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV	240
Qy	241	YGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFG	300

Db 241 YGLDQSKMNC DRVFN VFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFG 300

Qy 301 QKLNVCVSKQPAIMPGQSYGLE DGSCSYKDFSES RNNRFSTPEQAAKNR 349  
 |||

Db 301 QKLNVCVSKQPAIMPGQSYGLE DGSCSYKDFSES RNNRFSTPEQAAKNR 349

# RESULT 2

AAU27959

ID AAU27959 standard; protein; 589 AA.

XX

AC AAU27959;

XX

DT 18-DEC-2001 (first entry)

XX

DE Human contig polypeptide sequence #112.

XX

KW Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;  
 KW mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;  
 KW cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;  
 KW nervous system disorder; inflammatory disorder; cell differentiation;  
 KW angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;  
 KW genetic disorder; bone regeneration; tendon; ligament; tissue repair;  
 KW cytostatic; antirheumatic; antiarthritic; vulnerary; antiinflammatory;  
 KW antibacterial; immunosuppressive; vasotropic; antiparkinsonian;  
 KW neuroprotective; osteopathic; antidiabetic; antiasthmatic; antiallergic;  
 KW immunostimulant; analgesic; gene therapy.

XX

OS Homo sapiens.

OS Synthetic.

XX

PN WO200164834-A2.

XX

PD 07-SEP-2001.

XX

PF 26-FEB-2001; 2001WO-US004926.

XX

PR 28-FEB-2000; 2000US-00515126.

PR 18-MAY-2000; 2000US-00577409.

PR 17-JUN-2000; 2000US-00597707.

PR 14-JUL-2000; 2000US-00616807.

PR 19-SEP-2000; 2000US-00664641.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;  
 PI Drmanac R;

XX

DR WPI; 2001-589862/66.

DR N-PSDB; AAS44859.

XX

PT Novel polypeptides and nucleic acids obtained from cDNA libraries  
 PT prepared from various human tissues, for diagnosis, treatment of cancer,  
 PT neurological, inflammatory disorders and for use in arrays for detection.

XX

PS Claim 10; Page 137-138; 153pp; English.

Sequences AAU27676-AAU28019 represent full-length polypeptides and contiguous polypeptides of the invention. The proteins and their associated DNA sequences are useful for the treatment, diagnosis and prevention of various types of disorder in a mammalian subject such as a human, dog, monkey, mouse, hamster or rat. The disorders include cancers such as leukaemia, lymphoma and neuroblastoma, autoimmune disorders such as multiple sclerosis, connective tissue disease, rheumatoid arthritis, diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system disorders such as Parkinson's disease, Alzheimer's disease, Huntington's chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and Wernicke disease, inflammatory disorders such as nephritis, Crohn's disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory bowel disease. The sequences exhibit activity relating to angiogenesis, cell proliferation, cell differentiation, stem cell growth factor, activin or inhibin. Therefore, they can be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, accidental damage or genetic disorders. The sequences may also be used for regeneration of bone, cartilage, tendons and ligaments and in tissue repair and burn healing. Note: Some sequences for this patent did not form part of the printed specification, but were obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](ftp.wipo.int/pub/published_pct_sequences)

SO Sequence 589 AA;

Qy	1	VLGACNAVNYAADNQIYYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSI	60
Db	147	VLGACNAVNYAADNQIYYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSI	206
Qy	61	TTDVLYTICNPCGVPQQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE	120
Db	207	TTDVLYTICNPCGVPQQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE	266
Qy	121	YAKPTRLNVFKNDQDQDTWDYTNPNLSGQGDGPGSNPNKRQRQPPLLGDHPAEYGGPHGGYHS	180
Db	267	YAKPTRLNVFKNDQDQDTWDYTNPNLSGQGDGPGSNPNKRQRQPPLLGDHPAEYGGPHGGYHS	326
Qy	181	HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV	240
Db	327	HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV	386
Qy	241	YGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFG	300
Db	387	YGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFG	446
Qy	301	QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESRRNRFSTPEQAAKNR	349
Db	447	OKLNVCVSKOPAIMPGOSYGLEDGSCSYKDFSESRRNRFSTPEOAAKNR	495

RESULT 3  
AAU27787

ID AAU27787 standard; protein; 589 AA.  
XX  
AC AAU27787;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Human full-length polypeptide sequence #112.  
XX  
KW Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;  
KW mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;  
KW cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;  
KW nervous system disorder; inflammatory disorder; cell differentiation;  
KW angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;  
KW genetic disorder; bone regeneration; tendon; ligament; tissue repair;  
KW cytostatic; antirheumatic; antiarthritic; vulnerary; antiinflammatory;  
KW antibacterial; immunosuppressive; vasotropic; antiparkinsonian;  
KW neuroprotective; osteopathic; antidiabetic; antiasthmatic; antiallergic;  
KW immunostimulant; analgesic; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO200164834-A2.  
XX  
PD 07-SEP-2001.  
XX  
PF 26-FEB-2001; 2001WO-US004926.  
XX  
PR 28-FEB-2000; 2000US-00515126.  
PR 18-MAY-2000; 2000US-00577409.  
PR 17-JUN-2000; 2000US-00597707.  
PR 14-JUL-2000; 2000US-00616807.  
PR 19-SEP-2000; 2000US-00664641.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;  
PI Drmanac R;  
XX  
DR WPI; 2001-589862/66.  
DR N-PSDB; AAS44687.  
XX  
PT Novel polypeptides and nucleic acids obtained from cDNA libraries  
PT prepared from various human tissues, for diagnosis, treatment of cancer,  
PT neurological, inflammatory disorders and for use in arrays for detection.  
XX  
PS Claim 10; SEQ ID NO 284; 153pp; English.  
XX  
CC Sequences AAU27676-AAU28019 represent full-length polypeptides and contig  
CC polypeptides of the invention. The proteins and their associated DNA  
CC sequences are useful for the treatment, diagnosis and prevention of  
CC various types of disorder in a mammalian subject such as a human, dog,  
CC monkey, mouse, hamster or rat. The disorders include cancers such as  
CC leukaemia, lymphoma and neuroblastoma, autoimmune disorders such as  
CC multiple sclerosis, connective tissue disease, rheumatoid arthritis,  
CC diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system  
CC disorders such as Parkinson's disease, Alzheimer's disease, Huntington's

CC chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and  
 CC Wernicke disease, inflammatory disorders such as nephritis, Crohn's  
 CC disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory  
 CC bowel disease. The sequences exhibit activity relating to angiogenesis,  
 CC cell proliferation, cell differentiation, stem cell growth factor,  
 CC activin or inhibin. Therefore, they can be used to manipulate stem cells  
 CC in culture to give rise to neuroepithelial cells that can be used to  
 CC augment or replace cells damaged by illness, accidental damage or genetic  
 CC disorders. The sequences may also be used for regeneration of bone,  
 CC cartilage, tendons and ligaments and in tissue repair and burn healing.  
 CC Note: Some sequences for this patent did not form part of the printed  
 CC specification, but were obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 589 AA;

Query Match 100.0%; Score 1921; DB 4; Length 589;

Best Local Similarity 100.0%; Pred. No. 5.3e-169;

Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy      1 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSI 60
      |||
Db      147 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSI 206

Qy      61 TTDVLYTICNPGPVQRIVIFRKNVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 120
      |||
Db      207 TTDVLYTICNPGPVQRIVIFRKNVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 266

Qy      121 YAKPTRLNVFKNDQDTWDYTNPNLSGQDPSNPNKRQRQPPLLGDHPAEYGGPHGGYHS 180
      |||
Db      267 YAKPTRLNVFKNDQDTWDYTNPNLSGQDPSNPNKRQRQPPLLGDHPAEYGGPHGGYHS 326

Qy      181 HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV 240
      |||
Db      327 HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV 386

Qy      241 YGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNFMFG 300
      |||
Db      387 YGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNFMFG 446

Qy      301 QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESNNRFSTPEQAAKNR 349
      |||
Db      447 QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESNNRFSTPEQAAKNR 495

```

#### RESULT 4

ABO52974

ID ABO52974 standard; protein; 558 AA.

XX

AC ABO52974;

XX

DT 09-OCT-2003 (first entry)

XX

DE Human spliceosome associated protein (SAP) #91.

XX

KW Human; SAP; spliceosome associated protein; ribonucleoprotein;

KW RNP complex; RNA affinity substrate; RNP assembly sequence;

KW spliceosomal complex; hnRNP complex; mRNA export complex;  
 KW mRNA localisation complex; RNA editing complex; intron complex;  
 KW H complex; telomerase complex; fragile X protein complex;  
 KW reverse transcriptase complex; gene splicing complex.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2003068803-A1.  
 XX  
 PD 10-APR-2003.  
 XX  
 PF 14-JAN-2002; 2002US-00047991.  
 XX  
 PR 12-JAN-2001; 2001US-0261521P.  
 XX  
 PA (REED/) REED R.  
 PA (ZHOU/) ZHOU Z.  
 XX  
 PI Reed R, Zhou Z;  
 XX  
 DR WPI; 2003-540885/51.  
 XX  
 PT Isolating ribonucleoprotein complex, by contacting RNA affinity substrate  
 PT having ribonucleoprotein assembly sequence and affinity tag, with protein  
 PT mixture, subjecting complex formed to chromatography, affinity selection.  
 XX  
 PS Claim 24; Page; 39pp; English.  
 XX  
 CC The invention relates to forming (M1) an isolated ribonucleoprotein (RNP)  
 CC complex (C), involves contacting an RNA affinity substrate (S) comprising  
 CC an RNP assembly sequence (AS) and an affinity tag, with a protein mixture  
 CC to permit formation of (C) on AS, subjecting (C) to chromatographic  
 CC separation, and subjecting (C) to affinity selection, where the affinity  
 CC tag (e.g. bacteriophage MS2 coat protein in a fusion protein with E. coli  
 CC maltose binding protein) binds to an affinity matrix. Also included are  
 CC an isolated spliceosome preparation (isolated by (M1)), a RNA comprising  
 CC an RNP complex binding site and at least one phage coat protein  
 CC recognition site, a nucleic acid encoding the RNA, and treating (M2) a  
 CC subject having a disorder associated with abnormal RNP complexes (by  
 CC obtaining a sample of cells from a subject, purifying RNP complexes from  
 CC the cells of the subject by (M1), determining the presence in the  
 CC purified RNP complexes of one or more proteins, and normalising the  
 CC amount of RNPs in the subject. (M1) is useful for forming an isolated RNP  
 CC complex selected from a spliceosomal complex (selected from E, A, B and C  
 CC complex), an hnRNP complex, an mRNA export complex, an mRNA localisation  
 CC complex, an RNA editing complex, an intron complex, or an H complex. (M1)  
 CC is useful in a diagnostic assay for determining whether a subject has  
 CC abnormal RNP complexes, (M2) is useful for treating a subject having a  
 CC disorder associated with abnormal RNP complexes. (M1) is useful for  
 CC forming an isolated RNP complex such as a telomerase complex, a fragile X  
 CC protein complex, a reverse transcriptase complex or a gene splicing  
 CC complex. The present sequence represents a known human spliceosome  
 CC associated protein (SAP) isolated by the methods of the invention. Note:  
 CC The present sequence is not shown in the specification but was obtained  
 CC from Genbank or Swissprot using the information provided in table 1 of  
 CC the specification  
 XX



SQ Sequence 558 AA;

Query Match 99.4%; Score 1909; DB 6; Length 558;  
Best Local Similarity 99.7%; Pred. No. 6.4e-168;  
Matches 348; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSI 60
          |||
Db     116 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSI 175

Qy      61 TTDVLYTICNPGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 120
          |||
Db     176 TTDVLYTICNPGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 235

Qy     121 YAKPTRLNVFKNDQDTWDYTNPNLSGQGDPGSNPNKRQRPPLLDGHPAEYGGPHGGYHS 180
          |||
Db     236 YAKPTRLNVFKNDQDTWDYTNPNLSGQGDPGSNPNKRQRPPLLDGHPAEYGGPHGGYHS 295

Qy     181 HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV 240
          |||
Db     296 HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV 355

Qy     241 YGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNFMFG 300
          |||
Db     356 YGLDQSKMNGDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNFMFG 415

Qy     301 QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESNNRNFSTPEQAAKNR 349
          |||
Db     416 QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESNNRNFSTPEQAAKNR 464
```

RESULT 5

ABB97144

ID ABB97144 standard; protein; 589 AA.

XX

AC ABB97144;

XX

DT 21-JUN-2002 (first entry)

XX

DE Human tumour antigen related protein SEQ ID NO 46.

XX

KW Human; tumour; antigen; HLA-A2; cytotoxic T cell; cytostatic; cancer;  
KW vaccine.

XX

OS Homo sapiens.

XX

PN WO200210369-A1.

XX

PD 07-FEB-2002.

XX

PF 30-JUL-2001; 2001WO-JP006526.

XX

PR 31-JUL-2000; 2000JP-00231814.

XX

PA (ITOH/) ITOH K.

XX

PI Itoh K;

XX  
 DR WPI; 2002-291857/33.  
 DR N-PSDB; ABL56072.  
 XX  
 PT Tumor antigens inducing and/or activating HLA-A2-restricted tumor-  
 PT specific cytotoxic T cells, useful in diagnosis of and screening drugs  
 PT e.g. cancer vaccines for specific treatment of pancreatic cancer.  
 XX  
 PS Claim 2; Page 94-96; 127pp; Japanese.  
 XX  
 CC The invention relates to a peptide comprising an amino acid sequence  
 CC selected from 44 fully defined amino acid sequences (ABB96906-ABB969549)  
 CC and a polypeptide comprising an amino acid sequence selected from the 9  
 CC fully defined amino acid sequences (ABB97143-ABB97151). The above  
 CC comprise a tumour antigen inducing or activating HLA-A2-restricted tumour  
 CC -specific cytotoxic T cells, which recognise HLA-A2 and a tumour antigen  
 CC peptide and is thus activated. The peptides and polypeptides have  
 CC cytostatic activity. The tumour antigen is useful in diagnosis of and  
 CC screening drugs for specific treatment of pancreatic cancer, colon cancer  
 CC and stomach cancer including in the form of vaccines. The present  
 CC sequence is that of a tumour antigen protein, useful to the invention  
 XX  
 SQ Sequence 589 AA;

Query Match 99.4%; Score 1909; DB 5; Length 589;  
 Best Local Similarity 99.7%; Pred. No. 6.8e-168;  
 Matches 348; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSI	60
Db	147	VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSI	206
Qy	61	TTDVLYTICNPCGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE	120
Db	207	TTDVLYTICNPCGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE	266
Qy	121	YAKPTRLNVFKNDQDTWDYTNPNLSGQDPSNPNKRQRQPPLLGDHPAEYGGPHGGYHS	180
Db	267	YAKPTRLNVFKNDQDTWDYTNPNLSGQDPSNPNKRQRQPPLLGDHPAEYGGPHGGYHS	326
Qy	181	HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV	240
Db	327	HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV	386
Qy	241	YGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNFMFG	300
Db	387	YGLDQSKMNGDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNFMFG	446
Qy	301	QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESRRNRFSTPEQAAKNR	349
Db	447	QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESRRNRFSTPEQAAKNR	495

RESULT 6  
 ABG15420  
 ID ABG15420 standard; protein; 567 AA.  
 XX

AC ABG15420;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #15411.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS79607.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 20; SEQ ID NO 45779; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 567 AA;

Query Match 98.8%; Score 1897; DB 4; Length 567;  
Best Local Similarity 99.4%; Pred. No. 8.4e-167;  
Matches 347; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      1 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSI 60
      |||||
Db     125 VLGAGNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSI 184

Qy     61 TTDVLYTICNPCGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 120
      |||||
Db     185 TTDVLYTICNPCGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 244

Qy    121 YAKPTRLNVFKNDQDTWDYTNPNLSGQGDPSNPKNRQRPPLLGDHPAEYGGPHGGYHS 180
      |||||
Db    245 YAKPTRLNVFKNDQDTWDYTNPNLSGQGDPSNPKNRQRPPLLGDHPAEYGGPHGGYHS 304

Qy    181 HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV 240
      |||||
Db    305 HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV 364

Qy    241 YGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFG 300
      |||||
Db    365 YGLDQSKMNGDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFG 424

Qy    301 QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESNNRFSTPEQAAKNR 349
      |||||
Db    425 QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESNNRFSTPEQAAKNR 473
```

RESULT 7

AAU33004

ID AAU33004 standard; protein; 624 AA.

XX

AC AAU33004;

XX

DT 18-DEC-2001 (first entry)

XX

DE Novel human secreted protein #3495.

XX

KW Human; vaccination; gene therapy; nutritional supplement;

KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;

KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX

OS Homo sapiens.

XX

PN WO200179449-A2.

XX

PD 25-OCT-2001.

XX

PF 16-APR-2001; 2001WO-US008656.

XX

PR 18-APR-2000; 2000US-00552929.

PR 26-JAN-2001; 2001US-00770160.

XX

PA (HYSE-) HYSEQ INC.

XX



RESULT 8

AAU33002

ID AAU33002 standard; protein; 379 AA.

XX

AC AAU33002;

XX

DT 18-DEC-2001 (first entry)

XX

DE Novel human secreted protein #3493.

XX

KW Human; vaccination; gene therapy; nutritional supplement;

KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;

KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX

OS Homo sapiens.

XX

PN WO200179449-A2.

XX

PD 25-OCT-2001.

XX

PF 16-APR-2001; 2001WO-US008656.

XX

PR 18-APR-2000; 2000US-00552929.

PR 26-JAN-2001; 2001US-00770160.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Drmanac RT;

XX

DR WPI; 2001-611725/70.

XX

PT Nucleic acids encoding a range of human polypeptides, useful in genetic

PT vaccination, testing and therapy.

XX

PS Claim 20; Page 697; 765pp; English.

XX

CC The invention relates to novel human secreted polypeptides. The  
CC polypeptides and antibodies to the polypeptides are useful for  
CC determining the presence of or predisposition to a disease associated  
CC with altered levels of polypeptide. The polypeptides are also useful for  
CC identifying agents (agonists and antagonists) that bind to them. Cells  
CC expressing the proteins are useful for identifying a therapeutic agent  
CC for use in treatment of a pathology related to aberrant expression or  
CC physiological interactions of the polypeptide. Vectors comprising the  
CC nucleic acids encoding the polypeptides and cells genetically engineered  
CC to express them are also useful for producing the proteins. The proteins  
CC are useful in genetic vaccination, testing and therapy, and can be used  
CC as nutritional supplements. They may be used to increase stem cell  
CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon  
CC and/or nerve tissue growth or regeneration; immune suppression and/or  
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.  
CC AAU29510-AAU33304 represent the amino acid sequences of novel human  
CC secreted proteins of the invention

XX

SQ Sequence 379 AA;

Query Match 67.0%; Score 1286.5; DB 4; Length 379;  
 Best Local Similarity 64.5%; Pred. No. 2e-110;  
 Matches 254; Conservative 2; Mismatches 9; Indels 129; Gaps 5;

```

Qy      1 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSI 73

Qy     61 TTDVLYTICNPGPVQRIVIFRKNGVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 120
          ||
Db     74 TT----- 75

Qy    121 YAKPTRLNVFKNDQDTWDYTNPNLSGQGDPGSNPNKRQRQPPLLGDHPAEYGGPHGGYHS 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 ---PTRLNVFKNDQDTWDYTNPNLSGQGDPGSNPNKRQRQPPLLGDHPAEYGGPHGGYHS 132

Qy    181 HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV 240
          ||||||||||||||||||||| |||||||||:|
Db    133 HYHDEGYGPPPPHYEGRRMGPPVG-----EYGPHADSPVIMV 169

Qy    241 YGLDQSKMNCDRVFNVFCLYGNVEKVK--FMKSKPGAAMV--EMADGYAVDRAITHLNNN 296
          ||||||||||||||||||||| | || : | ||||||||||||||
Db    170 YGLDQSKMNCDRVFNVFCLYGNVEKVKISLKKQSPGGRPMGEEWLDGYAVDRAITHLNNN 229

Qy    297 FMFGQKLNVC-----VSKQPAIMP 315
          ||||||||| |||||||||
Db    230 FMFGQKLNVCVGAQAREGSRGTGERKGGEWGPAEEHSEAEVLTHTEMGCGSVSKQPAIMP 289

Qy    316 GQSYGLEDGSCSYKDFSESRRNRFSTPEQAAKNR 349
          ||||||||||||||||||||||||||||||
Db    290 GQSYGLEDGSCSYKDFSESRRNRFSTPEQAAKNR 323
  
```

RESULT 9

ABG15417

ID ABG15417 standard; protein; 404 AA.

XX

AC ABG15417;

XX

DT 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #15408.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US008631.

XX

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX

PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS79604.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX  
 PS Claim 20; SEQ ID NO 45776; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 404 AA;

Query Match 67.0%; Score 1286.5; DB 4; Length 404;  
 Best Local Similarity 64.5%; Pred. No. 2.2e-110;  
 Matches 254; Conservative 2; Mismatches 9; Indels 129; Gaps 5;

Qy 1 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGDSDDSRVNSVLLFTILNPIYSI 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 14 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGDSDDSRVNSVLLFTILNPIYSI 73  
 Qy 61 TTDVLYTICNPGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 120  
 ||  
 Db 74 TT----- 75  
 Qy 121 YAKPTRLNVFKNDQDQTDWDTNPNLSGQGDPSNPNKRQRQPPLLGDHPAEYGGPHGGYHS 180  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 76 ---PTRLNVFKNDQDQTDWDTNPNLSGQGDPSNPNKRQRQPPLLGDHPAEYGGPHGGYHS 132  
 Qy 181 HYHDEGYGPPPPHYEGRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV 240  
 ||||||||||||||||||||||||| |||||||||:  
 Db 133 HYHDEGYGPPPPHYEGRMGPPVG-----EYGPHADSPVIMV 169



Qy 241 YGLDQSKMNCDRVFNVFCLYGNVEKVK--FMKSKPGAAMV--EMADGYAVDRAITHLNNN 296  
 ||||||||||||||||||||| | || : | |||||||||||||  
 Db 170 YGLDQSKMNCDRVFNVFCLYGNVEKVKISLKKQSPGGRPMGEEWLDGYAVDRAITHLNNN 229

Qy 297 FMFGQKLNVC-----VSKQPAIMP 315  
 ||||||||| |||||||  
 Db 230 FMFGQKLNVCVGAQAREGSRGTGERKGGEWGPAEEHSEAEVLTHTEMGCGSVSKQPAIMP 289

Qy 316 GQSYGLEDGSCSYKDFSESRRNRFSTPEQAAKNR 349  
 |||||||||||||||||||||  
 Db 290 GQSYGLEDGSCSYKDFSESRRNRFSTPEQAAKNR 323

RESULT 10

ABP43680

ID ABP43680 standard; protein; 437 AA.

XX

AC ABP43680;

XX

DT 26-FEB-2003 (first entry)

XX

DE Human RNA associated protein 17.

XX

KW Neuroprotective; immunomodulator; cancer; cytostatic; anti-inflammatory;

KW gene therapy; nutritional supplement; wound; burn; ulcer;

KW Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;

KW autoimmune disorder; inflammation; vulnerary.

XX

OS Homo sapiens.

XX

PN WO200231111-A2.

XX

PD 18-APR-2002.

XX

PF 11-OCT-2001; 2001WO-US027760.

XX

PR 12-OCT-2000; 2000US-00687527.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;

PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;

XX

DR WPI; 2002-426278/45.

DR N-PSDB; ABQ60924.

XX

PT New polypeptides and their encoded proteins, useful as nutritional  
 PT sources or supplements, or in gene therapy, particularly for treating  
 PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or  
 PT inflammation.

XX

PS Claim 20; SEQ ID # 583; 357pp + Sequence Listing; English.

XX

CC The invention relates to 446 newly isolated polynucleotide sequences. The  
 CC activity of polynucleotides of the invention may be described as,  
 CC vulnerary, neuroprotective, immunomodulator, cytostatic and anti-



KW antimicrobial; cell proliferative disorder; inflammation; cirrhosis;  
 KW actinic keratosis; bursitis; arteriosclerosis; artherosclerosis;  
 KW hepatitis; myelofibrosis; primary thrombocythemia; psoriasis; cancer;  
 KW mixed connective tissue disease; MCTD; HIV; uveitis; Crohn's disease;  
 KW allergy; rheumatoid arthritis; parasitic infection.

XX

OS Homo sapiens.

XX

FH	Key	Location/Qualifiers
FT	Modified-site	6
FT		/note= "Potential phosphorylation site"
FT	Modified-site	30
FT		/note= "Potential phosphorylation site"
FT	Modified-site	41
FT		/note= "Potential phosphorylation site"
FT	Modified-site	56
FT		/note= "Potential phosphorylation site"
FT	Domain	73. .133
FT		/label= RNA_recognition_motif
FT	Modified-site	81
FT		/note= "Potential phosphorylation site"
FT	Modified-site	118
FT		/note= "Potential phosphorylation site"
FT	Modified-site	141
FT		/note= "Potential glycosylation site"
FT	Modified-site	144
FT		/note= "Potential phosphorylation site"
FT	Modified-site	145
FT		/note= "Potential phosphorylation site"
FT	Modified-site	149
FT		/note= "Potential phosphorylation site"
FT	Domain	166. .232
FT		/label= RNA_recognition_motif
FT	Modified-site	231
FT		/note= "Potential phosphorylation site"
FT	Modified-site	249
FT		/note= "Potential glycosylation site"
FT	Modified-site	254
FT		/note= "Potential phosphorylation site"
FT	Modified-site	280
FT		/note= "Potential phosphorylation site"
FT	Modified-site	312
FT		/note= "Potential phosphorylation site"
FT	Domain	332. .399
FT		/label= RNA_recognition_motif
FT	Modified-site	343
FT		/note= "Potential glycosylation site"
FT	Modified-site	421
FT		/note= "Potential phosphorylation site"
FT	Modified-site	488
FT		/note= "Potential phosphorylation site"
FT	Modified-site	520
FT		/note= "Potential glycosylation site"
FT	Modified-site	526
FT		/note= "Potential phosphorylation site"

XX

PN WO200011171-A2.

XX  
PD 02-MAR-2000.  
XX  
PF 20-AUG-1999; 99WO-US019361.  
XX  
PR 21-AUG-1998; 98US-0097550P.  
PR 12-JAN-1999; 99US-0115639P.  
XX  
PA (INCY-) INCYTE PHARM INC.  
XX  
PI Hillman JL, Yue H, Tang YT, Corley NC, Guegler KJ, Gorgone GA;  
PI Patterson C, Baughn MR, Lal P, Bandman O, Reddy R, Azimzai Y;  
PI Shih LL, Yang J, Lu DAM;  
XX  
DR WPI; 2000-237651/20.  
DR N-PSDB; AAZ51266.  
XX  
PT Human RNA-associated proteins useful in diagnosing, treating and  
PT preventing cell proliferative, autoimmune, inflammatory and infectious  
PT disorders.  
XX  
PS Claim 1; Page 96-97; 123pp; English.  
XX  
CC The present amino acid sequence is the human RNA-associated protein-17  
CC (RNAAP-17), identified in Incyte clone 2129080, derived from KIDNNOT05  
CC library. It is expressed in nervous, reproductive, gastrointestinal and  
CC haematopoietic/immune tissues. It has cytostatic, immunosuppressive,  
CC antiinflammatory, antiarteriosclerotic, hepatotropic, keratolytic,  
CC neuroprotective, antipsoriatic, anti-HIV, antiallergic, antirheumatic,  
CC virucide, antiarthritic, ophthalmological and antimicrobial activity.  
CC RNAAP antibodies are useful for diagnosis of diseases associated with  
CC altered expression or activity of RNAAP. It is used to treat cell  
CC proliferative, autoimmune, inflammatory and infectious disorders, like  
CC actinic keratosis, bursitis, arteriosclerosis, atherosclerosis,  
CC cirrhosis, hepatitis, myelofibrosis, mixed connective tissue disease  
CC (MCTD), psoriasis, primary thrombocythemia and cancer, HIV, allergies,  
CC rheumatoid arthritis, uveitis, Crohn's disease, and bacterial, viral and  
CC parasitic infections  
XX  
SQ Sequence 537 AA;

Query Match 50.8%; Score 976.5; DB 3; Length 537;  
Best Local Similarity 57.1%; Pred. No. 1.8e-81;  
Matches 198; Conservative 47; Mismatches 77; Indels 25; Gaps 7;

Qy 4 ACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSITTD 63  
| | :||| :|||| || |||||:::|::|::| | ||| :| ||:| || |  
Db 119 AKECVTFAADEPVYIAGQQAFFNYSTSKRITRPGNTDDPSGGNKVLLLSIQNPLYPITVD 178  
Qy 64 VLYTICNPCGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIEYAK 123  
||||:|||| | |||||:::|::|::|::| ||:||||:|||||:|||||:|  
Db 179 VLYTVCNPVGKVQRIVIFKRNGIQAMVEFESVLCAQKAKAALNGADIYAGCCTLKIEYAR 238  
Qy 124 PTRLNVFKNDQDTWDYTNPNLSGQDGPSPNPKRQRQPPLLGDHPAEYGGPHGGYHSHYH 183  
||||| :|| |:||| | | : | | |||| :||:|: : | || ||  
Db 239 PTRLNVIRNDNDSWDYTKPYL-GRRDRGKG---RQRQ-AILGEHPSSF--RHDGYGSH-- 289

Qy 184 DEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEY--GPHADSPVLMVY 241  
 || | || | | : | | | : ||  
 Db 290 ----GPLLPLPSRYRMG-----SRDTPELVAYPLPQASSSYMHGNGPSGSVVMVS 335

Qy 242 GLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNFMFGQ 301  
 || | ||| |||:|||||:|||||: || |:| | |||:|:|||| :||:  
 Db 336 GLHQLKMNC SRVFNLFCLYGNIEKVKFMKTIPGTALVEMGDEYAVERA VTHLNNVKLFGK 395

Qy 302 KLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESRRNRFSTPEQA AKN 348  
 :||||| :::| | : |||: |||: |:|:|: ||:|  
 Db 396 RLNVCVSKQH SVVPSQIFELDGTSSYKDFAMSKNNRFTSAGQASKN 442

RESULT 12

AAB41893

ID AAB41893 standard; protein; 537 AA.

XX

AC AAB41893;

XX

DT 08-FEB-2001 (first entry)

XX

DE Human ORFX ORF1657 polypeptide sequence SEQ ID NO:3314.

XX

KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive.

XX

OS Homo sapiens.

XX

PN WO200058473-A2.

XX

PD 05-OCT-2000.

XX

PF 31-MAR-2000; 2000WO-US008621.

XX

PR 31-MAR-1999; 99US-0127607P.

PR 02-APR-1999; 99US-0127636P.

PR 05-APR-1999; 99US-0127728P.

PR 30-MAR-2000; 2000US-00540763.

XX

PA (CURA-) CURAGEN CORP.

XX

PI Shimkets RA, Leach M;

XX

DR WPI; 2000-602362/57.

DR N-PSDB; AAC76102.

XX  
PT Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease.

XX  
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;  
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;  
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;  
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;  
CC antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The  
CC sequences can be used for determining the presence of or predisposition  
CC to, or preventing or treating pathological conditions associated with an  
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX  
CC proteins in gene therapy vectors. The proteins and nucleic acids may be  
CC used to treat cancers, proliferative disorders, neurodegenerative  
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester  
CC storage, systemic lupus erythematosus, severe combined immunodeficiency  
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to  
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive

Query Match 50.8%; Score 976.5; DB 3; Length 537;  
Best Local Similarity 57.1%; Pred. No. 1.8e-81;  
Matches 198; Conservative 47; Mismatches 77; Indels 25; Gaps 7;

RESULT 13

ADI63130

ID ADI63130 standard; protein; 542 AA.

XX

AC ADI63130;

XX

DT 22-APR-2004 (first entry)

XX

DE Human apoptosis-associated protein SEQ ID 573.

XX

KW apoptosis; cell death; cytostatic; neuroprotective; immunosuppressive;  
KW antirheumatic; antiarthritic; dermatological; antiinflammatory;  
KW hepatotropic; virucide; nootropic; anticonvulsant; antiparkinsonian;  
KW vasotropic; cerebroprotective; antialcoholic; gene therapy; tumour;  
KW autoimmune disease; degenerative disease; viral infection; leukaemia;  
KW carcinoma; sarcoma; multiple sclerosis; rheumatoid arthritis; diabetes;  
KW lupus; hepatitis; influenza viruses; Alzheimer's disease;  
KW Huntington's disease; Parkinson's diseases; reperfusion injury; stroke;  
KW alcoholic liver disease; human.

XX

OS Homo sapiens.

XX

PN WO2003058021-A2.

XX

PD 17-JUL-2003.

XX

PF 13-JAN-2003; 2003WO-EP000270.

XX

PR 11-JAN-2002; 2002DE-01000856.

XX

PA (XANT-) XANTOS BIOMEDICINE AG.

XX

PI Koenig-Hoffman K, Kazinski M, Schaefer R, Kesper B;

XX

DR WPI; 2003-542134/51.

XX

PT New nucleic acids involved in apoptosis, useful for diagnosis and  
PT treatment of e.g. tumors and degenerative disease, also related proteins,  
PT antibodies and modulators.

XX

PS Claim 1b; SEQ ID NO 573; 517pp; German.

XX

CC This invention describes novel nucleic acid molecules that are associated  
CC with apoptosis and encode a polypeptide and are derived from a normalised  
CC gene library (embryonic or liver) or clone collections, and the extent of  
CC apoptosis measured by cell death detection assay or the CPRG assay  
CC (measuring loss of membrane integrity). The products of the invention  
CC have cytostatic, neuroprotective, immunosuppressive, antirheumatic,  
CC antiarthritic, dermatological, antiinflammatory, hepatotropic, virucide,  
CC nootropic, anticonvulsant, antiparkinsonian, vasotropic,  
CC cerebroprotective and antialcoholic activity and can be used for gene  
CC therapy. The polynucleotides also related vectors, hosts (or their  
CC extracts), encoded polypeptide (or their receptors) and/or agents that  
CC inhibit their activity (including antisense sequences) are used for  
CC treatment or prevention of tumours, autoimmune or degenerative diseases





KW chemotaxis; food additive.  
XX  
OS Homo sapiens.  
XX  
PN WO200154472-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001307.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.

PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.

PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Rosen CA, Barash SC, Ruben SM;

XX

DR WPI; 2001-476159/51.

DR N-PSDB; ADM19525.

XX

PT Isolated nucleic acid molecule encoding a channel/transporter protein is  
PT used in preventing, treating or ameliorating a medical condition.

XX

PS Claim 11; SEQ ID NO 811; 809pp; English.

XX

CC The invention relates to an isolated nucleic acid molecule encoding a  
CC channel/transporter protein or sequences at least 95% identical to a  
CC these. The nucleic acids and proteins encoded by them are used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
CC in diagnosing a pathological condition or susceptibility to a  
CC pathological condition. The antibodies to the proteins can also be used  
CC in alleviating symptoms associated with the disorders and in diagnostic  
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays  
CC (ELISA). Disorders which are diagnosed or treated include autoimmune  
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.  
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac



KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;  
 KW immune disorder; haematopoietic cell disorder; autoimmune disorder;  
 KW allergic reaction; graft versus host disease; organ rejection;  
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;  
 KW neurological disease; drug screening.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200055350-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 08-MAR-2000; 2000WO-US005882.  
 XX  
 PR 12-MAR-1999; 99US-0124270P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM;  
 XX  
 DR WPI; 2000-587533/55.  
 DR N-PSDB; AAC78118.  
 XX  
 PT Novel isolated nucleic acids comprising sequences encoding peptides  
 PT useful for treating or diagnosing e.g. cancer.  
 XX  
 PS Claim 11; Page 2008-2009; 2352pp; English.  
 XX  
 CC AAC77607 to AAC78448 encode the human cancer associated proteins given in  
 CC AAB43398 to AAB44239. The proteins can have activities based on the  
 CC tissues and cells the genes are expressed in. Example of activities  
 CC include: cytostatic; proliferative; vulnerary; immunomodulator;  
 CC antidiabetic; antiasthmatic; antirheumatic; antiarthritic;  
 CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;  
 CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;  
 CC nootropic; vasotropic; antipsoriatic and antiangiogenic. The  
 CC polynucleotides and polypeptides can be used for preventing, treating or  
 CC ameliorating medical conditions and diagnosing pathological conditions.  
 CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from  
 CC the present invention may be used to treat immune disorders by activating  
 CC or inhibiting the proliferation, differentiation or mobilisation of  
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune  
 CC disorders, allergic reactions, graft versus host disease and organ  
 CC rejection, modulate haemostatic or thrombolytic activity, modulate  
 CC inflammation, cancers, cardiovascular disorders, neurological disease and  
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,  
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to  
 CC AAC78457 and AAB44240 represent sequences used in the exemplification of  
 CC the present invention  
 XX  
 SQ Sequence 301 AA;

Query Match 40.6%; Score 780; DB 3; Length 301;  
 Best Local Similarity 99.3%; Pred. No. 1.4e-63;  
 Matches 148; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGDSDDSRVNSVLLFTILNPIYSI 60

```

          |||||||
Db      137 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSI 196

Qy      61 TTDVLYTICNPGPVQRIVIFRKNGVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 120
          |||||||
Db      197 TTDVLYTICNPGPVQRIVIFRKNGVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 256

Qy      121 YAKPTRLNVFKNDQDTWDYTNPNSGQGD 149
          |||||||
Db      257 YAKPTRLNVFKNDQDTWDYTNPNSGQGN 285

```

Search completed: January 7, 2005, 14:48:24  
Job time : 71.1545 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2005, 13:52:30 ; Search time 22.8936 Seconds  
(without alignments)  
1010.981 Million cell updates/sec

Title: US-10-726-721A-7  
Perfect score: 1921  
Sequence: 1 VLGACNAVNYAADNQIYIAG.....DFSESRRNNRFSTPEQAANKR 349

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query Match	Length			
1	1921	100.0	349	4	US-09-780-996A-7	Sequence 7, Appli
2	401	20.9	77	1	US-07-881-075-8	Sequence 8, Appli
3	401	20.9	77	1	US-08-120-827-8	Sequence 8, Appli
4	401	20.9	77	1	US-08-478-675-8	Sequence 8, Appli
5	389	20.2	76	1	US-07-881-075-9	Sequence 9, Appli
6	389	20.2	76	1	US-08-120-827-9	Sequence 9, Appli
7	389	20.2	76	1	US-08-478-675-9	Sequence 9, Appli
8	341	17.8	450	4	US-09-270-767-57535	Sequence 57535, A
9	341	17.8	467	4	US-09-270-767-42256	Sequence 42256, A
10	169	8.8	76	1	US-07-881-075-7	Sequence 7, Appli
11	169	8.8	76	1	US-08-120-827-7	Sequence 7, Appli

12	169	8.8	76	1	US-08-478-675-7	Sequence 7, Appli
13	157	8.2	25	4	US-09-418-839-2	Sequence 2, Appli
14	152.5	7.9	78	1	US-07-881-075-5	Sequence 5, Appli
15	152.5	7.9	78	1	US-08-120-827-5	Sequence 5, Appli
16	152.5	7.9	78	1	US-08-478-675-5	Sequence 5, Appli
17	136	7.1	21	3	US-08-602-999A-427	Sequence 427, App
18	136	7.1	21	4	US-09-500-124-427	Sequence 427, App
19	134	7.0	523	4	US-09-248-796A-17555	Sequence 17555, A
20	134	7.0	684	4	US-09-823-240A-9	Sequence 9, Appli
21	133	6.9	276	4	US-09-538-092-889	Sequence 889, App
22	131	6.8	2142	4	US-09-538-092-1142	Sequence 1142, Ap
23	128.5	6.7	234	4	US-09-538-092-888	Sequence 888, App
24	126	6.6	76	1	US-07-881-075-6	Sequence 6, Appli
25	126	6.6	76	1	US-08-120-827-6	Sequence 6, Appli
26	126	6.6	76	1	US-08-478-675-6	Sequence 6, Appli
27	122.5	6.4	331	4	US-09-538-092-845	Sequence 845, App
28	122.5	6.4	424	4	US-09-538-092-1338	Sequence 1338, Ap
29	120	6.2	247	4	US-09-538-092-890	Sequence 890, App
30	119	6.2	251	4	US-09-538-092-840	Sequence 840, App
31	119	6.2	262	4	US-08-311-731A-333	Sequence 333, App
32	118.5	6.2	309	4	US-09-270-767-44846	Sequence 44846, A
33	118	6.1	391	4	US-09-538-092-1097	Sequence 1097, Ap
34	118	6.1	437	2	US-08-935-450-8	Sequence 8, Appli
35	118	6.1	437	4	US-09-338-123-8	Sequence 8, Appli
36	117.5	6.1	124	2	US-08-925-237-2	Sequence 2, Appli
37	117.5	6.1	1958	1	US-07-945-283-2	Sequence 2, Appli
38	117	6.1	617	1	US-08-137-614A-26	Sequence 26, Appl
39	117	6.1	637	3	US-08-072-064-1	Sequence 1, Appli
40	117	6.1	637	3	US-08-072-064-4	Sequence 4, Appli
41	117	6.1	637	3	US-08-072-064-6	Sequence 6, Appli
42	117	6.1	637	3	US-08-072-064-8	Sequence 8, Appli
43	117	6.1	637	5	PCT-US92-08558-1	Sequence 1, Appli
44	116.5	6.1	633	4	US-09-976-594-282	Sequence 282, App
45	116.5	6.1	633	4	US-09-821-687-10	Sequence 10, Appl

#### ALIGNMENTS

##### RESULT 1

US-09-780-996A-7

; Sequence 7, Application US/09780996A

; Patent No. 6696273

; GENERAL INFORMATION:

; APPLICANT: Maury, Isabella

; APPLICANT: Mercken, Luc

; APPLICANT: Fournier, Alain

; TITLE OF INVENTION: Partners of the PTB1 Domain of FE65, Preparation and Uses

; FILE REFERENCE: ST00004-US

; CURRENT APPLICATION NUMBER: US/09/780,996A

; CURRENT FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: FR00/01628

; PRIOR FILING DATE: 2000-02-10

; PRIOR APPLICATION NUMBER: US 60/198,500

; PRIOR FILING DATE: 2000-04-18

; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: PatentIn version 3.2



; SEQ ID NO 7  
; LENGTH: 349  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-780-996A-7

Query Match 100.0%; Score 1921; DB 4; Length 349;  
Best Local Similarity 100.0%; Pred. No. 1.1e-175;  
Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSI 60
        |||
Db      1 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSI 60

Qy     61 TTDVLYTICNPCGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 120
        |||
Db     61 TTDVLYTICNPCGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 120

Qy    121 YAKPTRLNVFKNDQDTWDYTNPNLSGQDPSNPNKRQRQPPLLGDHPAEYGGPHGGYHS 180
        |||
Db    121 YAKPTRLNVFKNDQDTWDYTNPNLSGQDPSNPNKRQRQPPLLGDHPAEYGGPHGGYHS 180

Qy    181 HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV 240
        |||
Db    181 HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV 240

Qy    241 YGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFG 300
        |||
Db    241 YGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFG 300

Qy    301 QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESNNRNFSTPEQAAKNR 349
        |||
Db    301 QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESNNRNFSTPEQAAKNR 349
```

RESULT 2

US-07-881-075-8

; Sequence 8, Application US/07881075  
; Patent No. 5444149  
; GENERAL INFORMATION:  
; APPLICANT: KEENE, JACK D.  
; APPLICANT: KING, PETER H.  
; APPLICANT: LEVINE, TODD  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE  
; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC  
ACIDS  
; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND  
IMMUNOREGULATION  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,  
; ADDRESSEE: P.C.  
; STREET: 1755 Jefferson Davis Highway, Fourth Floor  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/881,075
; FILING DATE: 19920511
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5444149man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 714-154-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 77 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-07-881-075-8

```

```

Query Match          20.9%; Score 401; DB 1; Length 77;
Best Local Similarity 100.0%; Pred. No. 5.1e-31;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      50 LFTILNPIYSITTDVLYTICNPGPVRIVIFRKNGVQAMVEFDSVQSAQRAKASLNGAD 109
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  LFTILNPIYSITTDVLYTICNPGPVRIVIFRKNGVQAMVEFDSVQSAQRAKASLNGAD 60

Qy     110 IYSGCCTLKIEYAKPTR 126
          ||||||||||||||||
Db     61 IYSGCCTLKIEYAKPTR 77

```

RESULT 3

US-08-120-827-8

```

; Sequence 8; Application US/08120827
; Patent No. 5525495

```

; GENERAL INFORMATION:

```

; APPLICANT: KEENE, JACK D.
; APPLICANT: KING, PETER H.
; APPLICANT: LEVINE, TODD

```

```

; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE
; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC
ACIDS

```

```

; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND
IMMUNOREGULATION

```

```

; NUMBER OF SEQUENCES: 101

```

; CORRESPONDENCE ADDRESS:

```

; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington

```

```

; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/120,827
; FILING DATE: 15-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5525495man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 714-158-0 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)413-3000
; TELEFAX: (703)413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 77 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-120-827-8

```

```

Query Match          20.9%; Score 401; DB 1; Length 77;
Best Local Similarity 100.0%; Pred. No. 5.1e-31;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      50 LFTILNPIYSITTDVLYTICNPGPVQRIVIFRKNGVQAMVEFDSVQSAQRAKASLNGAD 109
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 LFTILNPIYSITTDVLYTICNPGPVQRIVIFRKNGVQAMVEFDSVQSAQRAKASLNGAD 60

Qy     110 IYSGCCTLKIEYAKPTR 126
          ||||||||||||||||
Db      61 IYSGCCTLKIEYAKPTR 77

```

#### RESULT 4

US-08-478-675-8

; Sequence 8, Application US/08478675

; Patent No. 5773246

#### ; GENERAL INFORMATION:

; APPLICANT: KEENE, JACK D.

; APPLICANT: KING, PETER H.

; APPLICANT: LEVINE, TODD

; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE

; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS

; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION

; NUMBER OF SEQUENCES: 101

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,

```

; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,675
; FILING DATE: 07-JUN-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/120,827
; FILING DATE: 15-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5773246man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 714-158-0 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)413-3000
; TELEFAX: (703)413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 77 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-478-675-8

```

```

Query Match          20.9%; Score 401; DB 1; Length 77;
Best Local Similarity 100.0%; Pred. No. 5.1e-31;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      50 LFTILNPIYSITTDVLYTICNPGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGAD 109
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 LFTILNPIYSITTDVLYTICNPGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGAD 60

Qy      110 IYSGCCTLKIEYAKPTR 126
          |||||||||||||||
Db      61 IYSGCCTLKIEYAKPTR 77

```

RESULT 5

US-07-881-075-9

```

; Sequence 9, Application US/07881075
; Patent No. 5444149
; GENERAL INFORMATION:
; APPLICANT: KEENE, JACK D.
; APPLICANT: KING, PETER H.
; APPLICANT: LEVINE, TODD
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE

```

; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS  
 ; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION  
 ; NUMBER OF SEQUENCES: 51  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,  
 ; ADDRESSEE: P.C.  
 ; STREET: 1755 Jefferson Davis Highway, Fourth Floor  
 ; CITY: Arlington  
 ; STATE: Virginia  
 ; COUNTRY: U.S.A.  
 ; ZIP: 22202  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/881,075  
 ; FILING DATE: 19920511  
 ; CLASSIFICATION: 530  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Oblon, No. 5444149man F.  
 ; REGISTRATION NUMBER: 24,618  
 ; REFERENCE/DOCKET NUMBER: 714-154-0  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703)521-4500  
 ; TELEFAX: (703)486-2347  
 ; TELEX: 248855 OPAT UR  
 ; INFORMATION FOR SEQ ID NO: 9:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 76 amino acids  
 ; TYPE: AMINO ACID  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: peptide  
 US-07-881-075-9

Query Match 20.2%; Score 389; DB 1; Length 76;  
 Best Local Similarity 98.7%; Pred. No. 7e-30;  
 Matches 75; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 237 VLMVYGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNN 296  
 |||||||||||| ||||||||||||||||||||||||||||||||||||||||  
 Db 1 VLMVYGLDQSKMNGDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNN 60  
 Qy 297 FMFGQKLNVCVSKQPA 312  
 ||||||||||||  
 Db 61 FMFGQKLNVCVSKQPA 76

RESULT 6  
 US-08-120-827-9  
 ; Sequence 9, Application US/08120827  
 ; Patent No. 5525495  
 ; GENERAL INFORMATION:  
 ; APPLICANT: KEENE, JACK D.

```

; APPLICANT: KING, PETER H.
; APPLICANT: LEVINE, TODD
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE
; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC
ACIDS
; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND
IMMUNOREGULATION
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/120,827
; FILING DATE: 15-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5525495man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 714-158-0 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)413-3000
; TELEFAX: (703)413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 76 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-120-827-9

```

```

Query Match          20.2%; Score 389; DB 1; Length 76;
Best Local Similarity 98.7%; Pred. No. 7e-30;
Matches 75; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      237 VLMVYGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNN 296
          ||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db      1 VLMVYGLDQSKMNGDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNN 60

Qy      297 FMFGQKLNVCVSKQPA 312
          |||||||||||||||
Db      61 FMFGQKLNVCVSKQPA 76

```

```

RESULT 7
US-08-478-675-9
; Sequence 9, Application US/08478675

```

```

; Patent No. 5773246
; GENERAL INFORMATION:
;   APPLICANT:  KEENE, JACK D.
;   APPLICANT:  KING, PETER H.
;   APPLICANT:  LEVINE, TODD
;   TITLE OF INVENTION:  METHODS AND COMPOSITIONS USEFUL IN THE
;   TITLE OF INVENTION:  RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC
ACIDS
;   TITLE OF INVENTION:  INVOLVED IN CELL GROWTH, NEOPLASIA AND
IMMUNOREGULATION
;   NUMBER OF SEQUENCES:  101
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;     ADDRESSEE:  P.C.
;     STREET:    1755 Jefferson Davis Highway, Fourth Floor
;     CITY:      Arlington
;     STATE:     Virginia
;     COUNTRY:   U.S.A.
;     ZIP:       22202
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:     IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:     PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/478,675
;     FILING DATE:       07-JUN-1996
;     CLASSIFICATION:    536
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/120,827
;     FILING DATE:        15-SEP-1993
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Oblon, No. 5773246man F.
;     REGISTRATION NUMBER:  24,618
;     REFERENCE/DOCKET NUMBER:  714-158-0 CIP
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (703)413-3000
;     TELEFAX:    (703)413-2220
;     TELEX:      248855 OPAT UR
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  76 amino acids
;       TYPE:    amino acid
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  peptide
US-08-478-675-9

```

```

Query Match          20.2%;  Score 389;  DB 1;  Length 76;
Best Local Similarity 98.7%;  Pred. No. 7e-30;
Matches 75;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;

```

```

Qy      237 VLMVYGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNN 296
          ||||||||||| ||||||||||| ||||||||||| ||||||||||| |||||||||||
Db      1 VLMVYGLDQSKMNGDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNN 60

Qy      297 FMFGQKLNVCVSKQPA 312
          |||||||||||

```

## RESULT 8

US-09-270-767-57535

; Sequence 57535, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 57535

; LENGTH: 450

; TYPE: PRT

; ORGANISM: Drosophila melanogaster

US-09-270-767-57535

Query Match 17.8%; Score 341; DB 4; Length 450;

Best Local Similarity 28.2%; Pred. No. 3.3e-24;

Matches 107; Conservative 72; Mismatches 119; Indels 82; Gaps 16;

```

Qy      6 NAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSITTDVL 65
      | | ::| : | : : | :: : | :| : : :| :: | :|
Db      30 NNANSSSDS-----NSAMGILQNTSAVNAGGNTNAAGGPNTVLRVIVESLMYPVSLDIL 83

Qy      66 YTICNPGCPVQRIVIFRK-NGVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIEYAKP 124
      : | | | :| | | | | : :: : | | | : | :| :| :| :| :| :| :| :|
Db      84 HQIFQRYGKVLKIVTFTKNNSFQALIQYPDANSAQHAKSLLDGQNIYNGCCTLRIDNSKL 143

Qy      125 TRLNVFKNDQDTWDYTNPNLSGGQDPG-----SNPN-----KRQRQPPLLGDHP 168
      | | | | | : : | :| | | | | :| | | | | | | | | | | | | | | |
Db      144 TALNVKYNNDKSRDFTNPALP-PGEPGVDIMPTAGGLMNTNDLLLIAARQR-PSLSGDKI 201

Qy      169 AEYGGPHGGYHSHYHDEGYGPP---PPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPP--- 222
      | | | | | | : | : | | | | | | | | | | | | | | | | | | |
Db      202 V-----NGLGAPGVLPPFALG--LGTPLTG-----GYNNALPNLA 234

Qy      223 -----PPPPPEYGPHADSPVLMVYGLDQSKMNCDRVFNVFCLYGNVEKVKFM 269
      | | | | | | | | :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      235 AFSLANS GALQTTAPAMRGY-----SNVLLVSNLNEEMVTPDALFTLFGVYGDVQVRVKIL 289

Qy      270 KSKPGAAMVEMADGYAVDRAITHLNNNFMFGQKLNVCVSKQPAI-MPGQSYGLEDGSCSY 328
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      290 YNKKDSALIQMAEPQQAYLAMSHLDKRLRWGKPIRVMASKHQAVQLPKE--GQPDAGLT- 346

Qy      329 KDFSESRNNRFSTPEQAANK 348
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      347 RDYSQNPLHRFKKP--GSKN 364

```

## RESULT 9

US-09-270-767-42256

; Sequence 42256, Application US/09270767



```
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 42256
; LENGTH: 467
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-42256
```

```
Query Match          17.8%; Score 341; DB 4; Length 467;
Best Local Similarity 28.2%; Pred. No. 3.5e-24;
Matches 107; Conservative 72; Mismatches 119; Indels 82; Gaps 16;
```

```
Qy      6 NAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSITTDVL 65
      | | ::| :      : |      : :: |:: :      |::| : : :| :: |::
Db      47 NNANSSSDS-----NSAMGILQNTSAVNAGGNTNAAGGPNTVLRVIVESLMYPVSLDIL 100

Qy      66 YTICNPCGPVQRIVIFRK-NGVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIEYAKP 124
      : |      | | :|| | | | ||::: :      ||| ||: ||: |::|::|::|::|::|
Db      101 HQIFQRYGKVLKIVTFTTKNNSFQALIQYPDANSAQHAKSLLDGQNIYNGCCTLRIDNSKL 160

Qy      125 TRLNVFKNDQDTWDYTNPNLSGQGDPG-----SNPN-----KRQRQPPLLGDPH 168
      | ||| | : : |::| | | | :||      | |      ||| | | ||
Db      161 TALNVKYNNDKSRDFTNPALP-PGEPGVDIMPTAGGLMNTNDLLLLIAARQR-PSLSGDKI 218

Qy      169 AEYGGPHGGYHSHYHDEGYGPP---PPHYEGRRMGPPVGGHRRGSPRYGPQYGHPPP--- 222
      | | |      || | : | : |      | : |
Db      219 V-----NGLGAPGVLPPFALG--LGTPLTG-----GYNNALPNLA 251

Qy      223 -----PPPPPEYGPHADSPVLMVYGLDQSKMNCDRVFNVFCLYGNVEKVKFM 269
      |      |      | ||:| |:: : | :| :| :||::|::| :
Db      252 AFSLANS GALQTTAPAMRGY-----SNVLLVSNLNEEMVTPDALFTLFGVYGDVQRVKIL 306

Qy      270 KSKPGAAMVEMADGYAVDRAITHLNNNFMFGQKLNVCVSKQPAI-MPGQSYGLEDGSCSY 328
      :| :|::|::| :      |::|:: :|:: : | || |:: :| : | | :
Db      307 YNKKDSALIQMAEPQQAYLAMSHLDKLRLLWGKPIRVMASKHQAVQLPKE--GQPDAGLT- 363

Qy      329 KDFSESRRNRFSTPEQAAKN 348
      :|::|:: :|| | :||
Db      364 RDYSQNPLHRFKKP--GSKN 381
```

# RESULT 10

US-07-881-075-7

```
; Sequence 7, Application US/07881075
; Patent No. 5444149
; GENERAL INFORMATION:
; APPLICANT: KEENE, JACK D.
; APPLICANT: KING, PETER H.
; APPLICANT: LEVINE, TODD
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE
```

; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS  
 ; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION  
 ; NUMBER OF SEQUENCES: 51  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,  
 ; ADDRESSEE: P.C.  
 ; STREET: 1755 Jefferson Davis Highway, Fourth Floor  
 ; CITY: Arlington  
 ; STATE: Virginia  
 ; COUNTRY: U.S.A.  
 ; ZIP: 22202  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/881,075  
 ; FILING DATE: 19920511  
 ; CLASSIFICATION: 530  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Oblon, No. 5444149man F.  
 ; REGISTRATION NUMBER: 24,618  
 ; REFERENCE/DOCKET NUMBER: 714-154-0  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703)521-4500  
 ; TELEFAX: (703)486-2347  
 ; TELEX: 248855 OPAT UR  
 ; INFORMATION FOR SEQ ID NO: 7:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 76 amino acids  
 ; TYPE: AMINO ACID  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: peptide  
 US-07-881-075-7

Query Match 8.8%; Score 169; DB 1; Length 76;  
 Best Local Similarity 100.0%; Pred. No. 8.3e-09;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQK 32  
 ||||||||||||||||||||||||||||||||  
 Db 45 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQK 76

RESULT 11

US-08-120-827-7

; Sequence 7, Application US/08120827  
 ; Patent No. 5525495

; GENERAL INFORMATION:

; APPLICANT: KEENE, JACK D.  
 ; APPLICANT: KING, PETER H.  
 ; APPLICANT: LEVINE, TODD  
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE

; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS  
 ; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION  
 ; NUMBER OF SEQUENCES: 101  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,  
 ; ADDRESSEE: P.C.  
 ; STREET: 1755 Jefferson Davis Highway, Fourth Floor  
 ; CITY: Arlington  
 ; STATE: Virginia  
 ; COUNTRY: U.S.A.  
 ; ZIP: 22202  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/120,827  
 ; FILING DATE: 15-SEP-1993  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Oblon, No. 5525495man F.  
 ; REGISTRATION NUMBER: 24,618  
 ; REFERENCE/DOCKET NUMBER: 714-158-0 CIP  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703)413-3000  
 ; TELEFAX: (703)413-2220  
 ; TELEX: 248855 OPAT UR  
 ; INFORMATION FOR SEQ ID NO: 7:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 76 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: peptide  
 US-08-120-827-7

Query Match 8.8%; Score 169; DB 1; Length 76;  
 Best Local Similarity 100.0%; Pred. No. 8.3e-09;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQK 32  
 ||||||||||||||||||||||||||||||||  
 Db 45 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQK 76

RESULT 12

US-08-478-675-7

; Sequence 7, Application US/08478675

; Patent No. 5773246

; GENERAL INFORMATION:

; APPLICANT: KEENE, JACK D.

; APPLICANT: KING, PETER H.

; APPLICANT: LEVINE, TODD

; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE

; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS  
 ; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION  
 ; NUMBER OF SEQUENCES: 101  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,  
 ; ADDRESSEE: P.C.  
 ; STREET: 1755 Jefferson Davis Highway, Fourth Floor  
 ; CITY: Arlington  
 ; STATE: Virginia  
 ; COUNTRY: U.S.A.  
 ; ZIP: 22202  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/478,675  
 ; FILING DATE: 07-JUN-1996  
 ; CLASSIFICATION: 536  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/120,827  
 ; FILING DATE: 15-SEP-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Oblon, No. 5773246man F.  
 ; REGISTRATION NUMBER: 24,618  
 ; REFERENCE/DOCKET NUMBER: 714-158-0 CIP  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703)413-3000  
 ; TELEFAX: (703)413-2220  
 ; TELEX: 248855 OPAT UR  
 ; INFORMATION FOR SEQ ID NO: 7:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 76 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: peptide  
 US-08-478-675-7

Query Match 8.8%; Score 169; DB 1; Length 76;  
 Best Local Similarity 100.0%; Pred. No. 8.3e-09;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQK 32  
 |||||  
 Db 45 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQK 76

RESULT 13  
 US-09-418-839-2  
 ; Sequence 2, Application US/09418839  
 ; Patent No. 6617432  
 ; GENERAL INFORMATION:  
 ; APPLICANT: GETZENBERG, ROBERT H.  
 ; TITLE OF INVENTION: NUCLEAR MATRIX PROTEINS, POLYNUCLEOTIDE SEQUENCES

; TITLE OF INVENTION: ENCODING THEM, AND THEIR USE  
; FILE REFERENCE: 076333/0170  
; CURRENT APPLICATION NUMBER: US/09/418,839  
; CURRENT FILING DATE: 1999-10-15  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 25  
; TYPE: PRT  
; ORGANISM: Rattus sp.  
US-09-418-839-2

Query Match 8.2%; Score 157; DB 4; Length 25;  
Best Local Similarity 96.0%; Pred. No. 2.5e-08;  
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 213 YGPQYGHPPPPPPPEYGPHADSPV 237  
|||||||:|||||||  
Db 1 YGPQYGHPPPPPPPDYGPHADSPV 25

RESULT 14

US-07-881-075-5

; Sequence 5, Application US/07881075

; Patent No. 5444149

; GENERAL INFORMATION:

; APPLICANT: KEENE, JACK D.

; APPLICANT: KING, PETER H.

; APPLICANT: LEVINE, TODD

; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE

; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC  
ACIDS

; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND  
IMMUNOREGULATION

; NUMBER OF SEQUENCES: 51

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,

; ADDRESSEE: P.C.

; STREET: 1755 Jefferson Davis Highway, Fourth Floor

; CITY: Arlington

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/881,075

; FILING DATE: 19920511

; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:

; NAME: Oblon, No. 5444149man F.

; REGISTRATION NUMBER: 24,618

; REFERENCE/DOCKET NUMBER: 714-154-0

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703)521-4500  
; TELEFAX: (703)486-2347  
; TELEX: 248855 OPAT UR  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 78 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-07-881-075-5

Query Match 7.9%; Score 152.5; DB 1; Length 78;  
Best Local Similarity 44.4%; Pred. No. 3.3e-07;  
Matches 32; Conservative 15; Mismatches 24; Indels 1; Gaps 1;

Qy 55 NPIYSITTDVLYTICNPCGPVQRIVIFRKNG-VQAMVEFDSVQSAQRAKASLNGADIYSG 113  
| | : | | | | : | | : | | | | : : : | | | | | : | : | :  
Db 6 NLFYPVTLQVLMQIFSKFGTVLKIITFTKNNQFQALLQYADPVSAQHAKLSLDGQNIYNA 65  
  
Qy 114 CCTLKIEYAKPT 125  
| | | : | : : | |  
Db 66 CCTLRIDFSKLT 77

RESULT 15

US-08-120-827-5

; Sequence 5, Application US/08120827  
; Patent No. 5525495

; GENERAL INFORMATION:

; APPLICANT: KEENE, JACK D.  
; APPLICANT: KING, PETER H.  
; APPLICANT: LEVINE, TODD

; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE

; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC  
ACIDS

; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND  
IMMUNOREGULATION

; NUMBER OF SEQUENCES: 101

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,  
; ADDRESSEE: P.C.

; STREET: 1755 Jefferson Davis Highway, Fourth Floor

; CITY: Arlington

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/120,827

; FILING DATE: 15-SEP-1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Oblon, No. 5525495man F.

```

;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  714-158-0 CIP
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)413-3000
;      TELEFAX:  (703)413-2220
;      TELEX:  248855 OPAT UR
;      INFORMATION FOR SEQ ID NO:  5:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH:  78 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;      MOLECULE TYPE:  peptide
US-08-120-827-5

```

```

Query Match          7.9%;  Score 152.5;  DB 1;  Length 78;
Best Local Similarity 44.4%;  Pred. No. 3.3e-07;
Matches  32;  Conservative  15;  Mismatches  24;  Indels  1;  Gaps  1;

```

```

Qy      55 NPIYSITTDVLYTICNPGPVQRIVIFRKNG-VQAMVEFDSVQSAQRAKASLNGADIYSG 113
      | | :| || | : | | :| :| || ||::: || || ||:| :||:
Db      6  NLFYPVTLDVLMQIFSKFGTVLKIITFTKNNQFQALLQYADPVSAQHAKLSLDGQNIYNA 65

Qy      114 CCTLKIEYAKPT 125
      ||||:|:::| |
Db      66 CCTLRIDFSKLT 77

```

```

Search completed: January  7, 2005, 14:51:41
Job time : 23.8936 secs

```

OM protein - protein search, using sw model

Run on: January 7, 2005, 14:33:20 ; Search time 17.8061 Seconds  
 (without alignments)  
 1885.849 Million cell updates/sec

Title: US-10-726-721A-7  
 Perfect score: 1921  
 Sequence: 1 VLGACNAVNYAADNQIYIAG.....DFSESRRNRFSTPEQAAKNR 349

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : PIR\_79:\*  
 1: pir1:\*  
 2: pir2:\*  
 3: pir3:\*  
 4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1909	99.4	558	2	A33616	heterogeneous ribo
2	604.5	31.5	493	2	T15805	hypothetical prote
3	360	18.7	556	2	S36629	polypyrimidine tra
4	353	18.4	557	2	S26294	polypyrimidine tra
5	349.5	18.2	550	2	S23016	polypyrimidine tra
6	349	18.2	530	2	S15552	polypyrimidine tra
7	345	18.0	557	2	S68857	polypyrimidine tra
8	343.5	17.9	532	2	JC7526	polypyrimidine tra
9	327	17.0	528	2	A41718	polypyrimidine tra
10	296.5	15.4	584	2	A88299	protein D2089.4 [i
11	296.5	15.4	592	2	T20381	hypothetical prote
12	217.5	11.3	418	2	T51814	polypyrimidine tra
13	154	8.0	463	2	T10015	hypothetical prote



14	154	8.0	488	2	F86911	conserved hypothet
15	152.5	7.9	1621	2	T15264	hypothetical prote
16	150.5	7.8	250	1	S59118	small nuclear ribo
17	146	7.6	639	2	G02919	transcription fact
18	143	7.4	260	2	S22373	proline-rich prote
19	143	7.4	548	2	S52735	CW17R protein - mo
20	141.5	7.4	366	2	T26449	hypothetical prote
21	140	7.3	206	1	PIRT3	acidic proline-ric
22	139	7.2	166	1	PIHUSC	salivary proline-r
23	139	7.2	166	2	B25372	salivary proline-r
24	139	7.2	171	2	A27307	proline-rich phosph
25	137.5	7.2	2715	2	T13049	eyelid - fruit fly
26	136	7.1	148	2	S39206	proline-rich prote
27	134	7.0	253	2	S59117	small nuclear ribo
28	134	7.0	325	2	D70728	hypothetical prote
29	134	7.0	684	2	A56154	Abl substrate ena
30	133.5	6.9	170	2	A48013	proline-rich prote
31	133.5	6.9	471	2	T33997	hypothetical prote
32	133	6.9	310	1	PIHUSD	salivary proline-r
33	131	6.8	301	2	E29149	proline-rich prote
34	131	6.8	1870	2	S37671	MHC class III hist
35	131	6.8	1872	2	S36152	MHC class III hist
36	131	6.8	2142	2	B35098	MHC class III hist
37	130.5	6.8	412	2	B44418	surface antigen -
38	129.5	6.7	257	2	T10586	small nuclear ribo
39	129.5	6.7	273	2	C70551	hypothetical prote
40	129.5	6.7	414	2	JN0866	nucleolar protein
41	129.5	6.7	1776	2	G86280	protein T5E21.13 [
42	129	6.7	300	2	S19560	proline-rich prote
43	128.5	6.7	245	1	W4WL5	E4 protein - human
44	128.5	6.7	748	2	T04011	hypothetical prote
45	127.5	6.6	198	2	E86261	F13K23.6 protein -

#### ALIGNMENTS

##### RESULT 1

A33616

heterogeneous ribonuclear particle protein L - human

C;Species: Homo sapiens (man)

C;Date: 30-Mar-1990 #sequence\_revision 30-Mar-1990 #text\_change 09-Jul-2004

C;Accession: A33616

R;Pinol-Roma, S.; Swanson, M.S.; Gall, J.G.; Dreyfuss, G.

J. Cell Biol. 109, 2575-2587, 1989

A;Title: A novel heterogeneous nuclear RNP protein with a unique distribution on nascent transcripts.

A;Reference number: A33616; MUID:90078296; PMID:2687284

A;Accession: A33616

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-558 <PIN>

A;Cross-references: UNIPROT:P14866; GB:X16135; NID:g32355; PIDN:CAA34261.1; PID:g32356

C;Superfamily: Caenorhabditis elegans hypothetical protein C44B7.2

Query Match

99.4%; Score 1909; DB 2; Length 558;

Best Local Similarity 99.7%; Pred. No. 7.6e-141;  
Matches 348; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSI 60
      |||
Db     116 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSI 175

Qy      61 TTDVLYTICNPGCPVQRIVIFRKNGVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 120
      |||
Db     176 TTDVLYTICNPGCPVQRIVIFRKNGVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 235

Qy     121 YAKPTRLNVFKNDQDTWDYTNPNLSGQGDPGSNPNKRQRQPPLLGDHHPAEYGGPHGGYHS 180
      |||
Db     236 YAKPTRLNVFKNDQDTWDYTNPNLSGQGDPGSNPNKRQRQPPLLGDHHPAEYGGPHGGYHS 295

Qy     181 HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV 240
      |||
Db     296 HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV 355

Qy     241 YGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFG 300
      |||
Db     356 YGLDQSKMNGDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFG 415

Qy     301 QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESNNRNFSTPEQAAKNR 349
      |||
Db     416 QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESNNRNFSTPEQAAKNR 464
```

## RESULT 2

T15805

hypothetical protein C44B7.2 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 15-Sep-2000

C;Accession: T15805

R;Du, Z.

submitted to the EMBL Data Library, June 1995

A;Description: The sequence of *C. elegans* cosmid C44B7.

A;Reference number: S61146

A;Accession: T15805

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-493 <DUZ>

A;Cross-references: EMBL:U28928; NID:g861301; PID:g861311; PIDN:AAA68343.1;  
CESP:C44B7.2

A;Experimental source: strain Bristol N2

C;Genetics:

A;Gene: CESP:C44B7.2

A;Introns: 13/2; 45/3; 100/3; 201/3; 222/1; 289/3; 320/3

C;Superfamily: *Caenorhabditis elegans* hypothetical protein C44B7.2

Query Match 31.5%; Score 604.5; DB 2; Length 493;  
Best Local Similarity 40.2%; Pred. No. 1.9e-39;  
Matches 145; Conservative 49; Mismatches 116; Indels 51; Gaps 12;

```
Qy      3 GACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSITT 62
      || ||:| ||| : | | ||||| | | : | | ||: || | |
Db     82 GAKACVNFATSNQINVGGQGALFNYSTSQCIERMG--FESATPNKVLVTVLNAQYPIDA 139
```

```

Qy      63 DVLYTICNPCGPVQRIVIFRK-NGVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIEY 121
      ||:| | | | | |: : | ||:||||:|:: |: || ::||| ||||| |||:|:
Db      140 DVIYQISNAQGKVLRVAVMHKPTVVQALVEFESMEVAKAAKHAMNGADIYSGCCTLKVEF 199

Qy      122 AKPTRLNLFKNDQDQDTWDYTNP-NLSGQGDPGSNPNKRQRQPPLLGDHHPAIEYG-GPHGGYH 179
      ||| |: | : |:| |:| | | | : : : | | : || | | |
Db      200 AKPDRVRVQRQDKDQRDFTLPDNRNRPYEDDRNHYDRHDYQA-----PSSYGYSSRGGGH 253

Qy      180 SHYHDEGYGPPPPHYEGRRMGPPVGGHRRGSPRYGPQY---GHPPPPPPPEYGPHADSP 236
      | | | | | | | | | | | | | | | | | | | | | |
Db      254 SDY-----YGGDRGGPP----HPPPSRYRDDYEDRGYAQPAGGGP-----GC 291

Qy      237 VLMVYGLDQSKMNCDRVFNVFCLYGNVEKVKFKMKSCKPGAAMVEMADGYAVDRAITHLNNN 296
      |:|:| |: |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db      292 VMMIYGLEHGKINCDMLFNILCQYGNVLRISFMRTKTETGIIELGTPEERQNVLDLQGS 351

Qy      297 FMFGQKLVN-----CVS--KQPAIMPQGSYGLEDGSCSYKDFSESNNRNFSTPEQAAKN 348
      :|| | | | | | | | | | | | | | | | | | | | | |
Db      352 ALFGLTLEFKPSHQECVHHLRDPFLLP-----DGSPSKDYSSSRNQRFSTPELA 404

Qy      349 R 349
      |
Db      405 R 405

```

# RESULT 3

S36629

polypyrimidine tract-binding protein PTB-2 - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 06-Jan-1995 #sequence\_revision 01-Dec-2000 #text\_change 09-Jul-2004

C;Accession: S36629; S18669; S15553

R;Sengupta, P.

submitted to the EMBL Data Library, August 1993

A;Description: A rat myoblast protein recognizing DNA sequences in the 3'UTR of pro Alpha1(CI) collagen gene is a member of the family of .

A;Reference number: S36629

A;Accession: S36629

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-556 <SEN>

A;Cross-references: UNIPROT:Q00438; EMBL:X74565; NID:g397523; PIDN:CAA52653.1; PID:g397524

R;Brunel, F.; Alzari, P.M.; Ferrara, P.; Zakin, M.M.

Nucleic Acids Res. 19, 5237-5245, 1991

A;Title: Cloning and sequencing of PYBP, a pyrimidine-rich specific single strand DNA-binding protein.

A;Reference number: S18668; MUID:92020211; PMID:1681508

A;Accession: S18669

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 189-310, 'VPSHLCHPSR', 322-556 <BRU>

A;Cross-references: EMBL:X60790; NID:g57003; PIDN:CAA43203.1; PID:g57004

A;Note: submitted to the EMBL Data Library, July 1991

F;363-426/Domain: ribonucleoprotein repeat homology <RRM2>

Query Match

18.7%; Score 360; DB 2; Length 556;

Best Local Similarity 28.2%; Pred. No. 2.3e-20;

Matches 112; Conservative 63; Mismatches 130; Indels 92; Gaps 14;

```
Qy      8 VN YAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSR-----SVNS----- 47
      |||      : | | :: :|   ::      : :|      : |||
Db     110 VNYYTSVAPVLRGQPIYIQFSNHKELKTDSSPNQARAQAALQAVNSVQSGNLALRASAAA 169

Qy     48 -----VLLFTILNPIYSITTDVLYTICNPGCPVQRIVIFRKNQ-VQAMVEFDS 94
      ||      : | | :| ||| : | : | | :| :| | ||| :| ::|
Db     170 VDAGMAMAGQSPVLRRIIVENLFYPVTLVDLHQLFISKFQTVLKIITFTKNNQFQALLQYAD 229

Qy     95 VQSAQRAKASLNGADIYSGCCTLKIEYAKPTRLNVFKNDQDTWDYTNPNLSGQGDPGSNP 154
      ||| || ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db     230 PVSAQHAKLSLDGQNIYNACCTLRIDFSKLTSLNVKYNNDKSRDYTRPDLP-SGD----- 283

Qy    155 NKRQRQPPLLGDHPAEYGGPHGGYHSHYHDEGYGPPP-----PHYEGRRMGP- 201
      || |      | :| | | | | | | | | | | | | | | | | | | | | |
Db     284 ----SQPSLDQTMAAAFGAP--GIMSASPYAGAGFPPTFAIPQAAGLSVPNVHG-ALAPL 336

Qy    202 -----PVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMVYGLDQSKMNCDR 252
      | | |      | | |      | | |      | : ||| : | : :
Db     337 AIPSAAAAAAAAGRIAPGLAG-----AGNSVLLVSNLNPERVTPQS 378

Qy    253 VFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFQKLNVCVSKQPA 312
      : | : | : ||| : : || : : | | : ||| || | : ||| : : | : : : || :
Db     379 LFILFGVYGDVQRVKILFNKKENALVEMADGSQAQLAMSHLNGHKLHGKSVRITLSKHQS 438

Qy    313 I-MPGQSYGLEDGSCSYKDFSESRRNRFSTPEQAANK 348
      : : | : | || : ||| : | | : || | : ||
Db     439 VQLPRE--GQEDQGLT-KDYGSSPLHRFKKP--GSKN 470
```

#### RESULT 4

S26294

polypyrimidine tract-binding protein PTB-1 [validated] - human

N;Alternate names: 57k RNA-binding protein pPTB-1; heterogenous nuclear ribonucleoprotein I; heterogenous ribonuclear particle protein I; polypyrimidine tract-binding protein PTB-4

C;Species: Homo sapiens (man)

C;Date: 25-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 09-Jul-2004

C;Accession: S26294; S23017; A40325; A40324; B60472; S16046; S23015

R;Ghetti, A.; Pinol-Roma, S.; Michael, W.M.; Morandi, C.; Dreyfuss, G.

Nucleic Acids Res. 20, 3671-3678, 1992

A;Title: hnRNP I, the polypyrimidine tract-binding protein: distinct nuclear localization and association with hnRNAs.

A;Reference number: S26294; MUID:92350668; PMID:1641332

A;Accession: S26294

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-557 <GHE>

A;Cross-references: UNIPROT:Q9BUQ0; EMBL:X66975; NID:g32353; PIDN:CAA47386.1; PID:g32354

R;Patton, J.G.

submitted to the EMBL Data Library, May 1992

A;Reference number: S23016

A;Accession: S23017

A;Status: preliminary

A;Molecule type: DNA  
 A;Residues: 1-557 <PAT1>  
 A;Cross-references: EMBL:X65372; NID:g35771; PIDN:CAA46444.1; PID:g35772  
 R;Patton, J.G.; Mayer, S.A.; Tempst, P.; Nadal-Ginard, B.  
 Genes Dev. 5, 1237-1251, 1991  
 A;Title: Characterization and molecular cloning of polypyrimidine tract-binding protein: a component of a complex necessary for pre-mRNA splicing.  
 A;Reference number: A40325; MUID:91293584; PMID:1906036  
 A;Accession: A40325  
 A;Status: not compared with conceptual translation  
 A;Molecule type: mRNA  
 A;Residues: 1-298,325-557 <PAT2>  
 A;Cross-references: GB:X62006; NID:g35767; PIDN:CAA43973.1; PID:g35768  
 A;Note: part of this sequence was confirmed by protein sequencing  
 R;Gil, A.; Sharp, P.A.; Jamison, S.F.; Garcia-Blanco, M.A.  
 Genes Dev. 5, 1224-1236, 1991  
 A;Title: Characterization of cDNAs encoding the polypyrimidine tract-binding protein.  
 A;Reference number: A40324; MUID:91293583; PMID:1906035  
 A;Accession: A40324  
 A;Molecule type: mRNA  
 A;Residues: 1-298,325-557 <GIL>  
 A;Cross-references: EMBL:X60648; NID:g35773; PIDN:CAA43056.1; PID:g35774  
 A;Note: part of this sequence was confirmed by protein sequencing  
 R;Wittwer, C.U.; Bauw, G.; Krokan, H.E.  
 Biochemistry 28, 780-784, 1989  
 A;Title: Purification and determination of the NH-2-terminal amino acid sequence of uracil-DNA glycosylase from human placenta.  
 A;Reference number: A60472; MUID:89229080; PMID:2713345  
 A;Accession: B60472  
 A;Molecule type: protein  
 A;Residues: 353-367,'X',369-373,'X',375-376,'N',378 <WIT>  
 A;Note: this protein was sequenced after co-purification with uracil-DNA glycosylase from human placenta. Tentative identifications were made for six of the last eight residues  
 C;Comment: This protein binds to the polypyrimidine tract of mammalian introns.  
 C;Genetics:  
 A;Gene: GDB:PTB; PTB-1  
 A;Cross-references: GDB:132677  
 A;Map position: 14q23-14q24.1  
 C;Keywords: alternative splicing; splicing protein

Query Match 18.4%; Score 353; DB 2; Length 557;  
 Best Local Similarity 28.2%; Pred. No. 8.1e-20;  
 Matches 112; Conservative 62; Mismatches 131; Indels 92; Gaps 14;

```

Qy      8 VNYAADNQIYIAGHPAFVNYSTSQKISRPGSDDDSR-----SVNS----- 47
      |||      : | | : : |  ::      : :|      :|||
Db     111 VNYTTSVTPVLRGQPIYIQFSNHKELKTDSSPNQARAQAALQAVNSVQSGNLALAASAAA 170

Qy     48 -----VLLFTILNPIYSITTDVLYTICNPGVPVQRIVIFRKNG-VQAMVEFDS 94
      ||      : | | : | ||| : | : | | :| | | ||| :| :|
Db     171 VDAGMAMAGQSPVLRRIIVENLFYPVTLVDVLHQIFSKFGTVLKIITFTKNNQFQALLQYAD 230

Qy     95 VQSAQRAKASLNGADIYSGCCTLKIEYAKPTRLNVFKNDQDTWDYTNPNLSGQGDPGSNP 154
      ||| || ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db     231 PVSAQHAKLSLDGQNIYNACCTLRIDFSKLTSLNVKYNNDKSRDYTRPDLP-SGD----- 284
  
```

```

Qy      155 NKRQRQPPLLGDHPAEYGGPHGGYHSHYHDEGYGPPP-----PHYEGRRMGP- 201
          || |      | : | | |      | | ||      | : | : |
Db      285 ----SQPSLDQTMAAAFGAP--GIISASPYAGAGFPPTFAIPQAAGLSVPNVHG-ALAPL 337

Qy      202 -----PVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMVYGLDQSKMNCDR 252
          |      | |      |      | : ||:| | : ::
Db      338 AIPSAIAAAGRIAIPLAG-----AGNSVLLVSNLNPVRTPQS 379

Qy      253 VFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMMFGQKLNVCVSK-QP 311
          :| :| :||:|:| : :| | :|:| | | :| :| : :| |
Db      380 LFILFGVYGDVQVRKILFNKKENALVQMDGNQAQLAMSHLNGHKLHGKPIRITLSKHQN 439

Qy      312 AIMPGQSYGLEDGSCSYKDFSESRRNRFSTPEQAANK 348
          :| : | || : ||: | :|| | :||
Db      440 VQLPRE--GQEDQGLT-KDYGNSPLHRFKKP--GSKN 471

```

RESULT 5

S23016

polypyrimidine tract-binding protein PTB-2 - human

C;Species: Homo sapiens (man)

C;Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Jul-2004

C;Accession: S23016

R;Patton, J.G.

submitted to the EMBL Data Library, May 1992

A;Reference number: S23016

A;Accession: S23016

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-550 <PAT>

A;Cross-references: UNIPROT:P26599; EMBL:X65371; NID:g35769; PIDN:CAA46443.1;

PID:g35770

```

Query Match          18.2%; Score 349.5; DB 2; Length 550;
Best Local Similarity 27.5%; Pred. No. 1.5e-19;
Matches 109; Conservative 63; Mismatches 126; Indels 99; Gaps 14;

```

```

Qy      8 VNAAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSR-----SVNS----- 47
          |||      : | | :: :| ::      : :|      :||
Db      111 VNYYTSVTPVLRGQPIYIQFSNHKELKTDSSPNQARAQAALQAVNSVQSGNLALAASAAA 170

Qy      48 -----VLLFTILNPIYSITTDVLYTICNPGCPVQRIVIFRKNQ-VQAMVEFDS 94
          || : | | :| ||: | : | | :| :| | | :| :| :|
Db      171 VDAGMAMAGQSPVLRRIIVENLFYPVTLDVLHQIFSKFGTVLKIITFTKNNQFQALLQYAD 230

Qy      95 VQSAQRAKASLNGADIYSGCCTLKIEYAKPTRLNVFKNDQDQTDWDTNPNLSGQGDPSNP 154
          ||| || ||:| :||: ||||:|:| | || | : : ||| |:| ||
Db      231 PVSAQHAKLSLDGQNIYNACCTLRIDFSKLTSLNVKYNNDKSRDYTRPDLP-SGD----- 284

Qy      155 NKRQRQPPLLGDHPAEYGGPHGGYHSHYHDEGYGPPP-----PHYEGRRMGP- 201
          || |      | : | :      | | ||      | : | : |
Db      285 ----SQPSLDQTMAAAFASPYA-----GAGFPPTFAIPQAAGLSVPNVHG-ALAPL 330

Qy      202 -----PVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMVYGLDQSKMNCDR 252
          |      | |      |      | : ||:| | : ::
Db      331 AIPSAIAAAGRIAIPLAG-----AGNSVLLVSNLNPVRTPQS 372

```

```

Qy      253 VFNVFLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFGQKLNVCVSK-QP 311
      :| :| :||:|:| :| :| :|:| | | :| | | : :| : : :| |
Db      373 LFILFGVYGDVQVRVKILFNKKENALVQMADGNQAQLAMSHLNHKLHGKPIRITLSKHQN 432

Qy      312 AIMPGQSYGLEDGSCSYKDFSESRRNRFSTPEQAANK 348
      :| :| | | :| :| :| | :| | :| |
Db      433 VQLPRE--GQEDQGLT-KDYGNSPLHRFKKP--GSKN 464

```

# RESULT 6

S15552

polypyrimidine tract-binding protein 1 - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Jul-2004

C;Accession: S15552; S18668

R;Brunel, F.; Alzari, P.; Ferrara, P.; Zakin, M.M.

submitted to the EMBL Data Library, July 1991

A;Reference number: S15552

A;Accession: S15552

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-530 <BRU>

A;Cross-references: UNIPROT:Q00438; EMBL:X60789; NID:g57001; PIDN:CAA43202.1; PID:g57002

R;Brunel, F.; Alzari, P.M.; Ferrara, P.; Zakin, M.M.

Nucleic Acids Res. 19, 5237-5245, 1991

A;Title: Cloning and sequencing of PYBP, a pyrimidine-rich specific single strand DNA-binding protein.

A;Reference number: S18668; MUID:92020211; PMID:1681508

A;Accession: S18668

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-530 <BRU2>

A;Cross-references: EMBL:X60789; NID:g57001; PIDN:CAA43202.1; PID:g57002

F;337-400/Domain: ribonucleoprotein repeat homology <RRM2>

Query Match 18.2%; Score 349; DB 2; Length 530;

Best Local Similarity 27.1%; Pred. No. 1.6e-19;

Matches 105; Conservative 61; Mismatches 123; Indels 98; Gaps 11;

```

Qy      8 VN YAADNQIYIAGHPAFVNYSTSQKISRPGSDDDSR-----SVNS----- 47
      ||| :| | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      110 VNYYTSVAPVLRGQPIYIQFSNHKELKTDSSPNQARAQAALQAVNSVQSGNLALAASAAA 169

Qy      48 -----VLLFTILNPIYSITTDVLYTICNPCGPVQRIVIFRKNQ-VQAMVEFDS 94
      || :| | :| ||| :| :| | :| :| :| :| :| :| :| :| :| :|
Db      170 VDAGMAMAGQSPVLRRIIVENLFYPVTLDVLHQIFSKFGTVLKIITFTKNNQFQALLQYAD 229

Qy      95 VQSAQRAKASLNGADIYSGCCTLKIEYAKPTRLNVFKNDQDTWDYTNPNL-SGQGDPG-- 151
      ||| || ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      230 PVSAQHAKLSLDGQNIYNACCTLRIDFSKLTSLNVKYNNDKSRDYTRPDLP SGDSQPSLD 289

Qy      152 -----SNPNKRQRPPLLDGHPAEYGGPHGGYHSHYHDEGYGPPPPHYEGRMGPP 202
      || | | | | | | | | | | | | | | | | | | | | | | | |
Db      290 QTMAAAFGLSVPNVHGALAPLAIPSAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIA 329

```

```

Qy      203 VGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMVYGLDQSKMNCDRVFNVFCLYGN 262
      : |                               | : ||:| |: ::      :| :| :||:
Db      330 LAG-----AGNSVLLVSNLNPERVTPQSLFILFGVYGD 362

Qy      263 VEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFGQKLNVCVSKQPAI-MPGQSYGL 321
      |::|| : :| .:||||| |::||| : : |: : : :|| :: :| : |
Db      363 VQRVKILFNKKENALVEMADGSQAQLAMSHLNGHKLHGKSVRITLSKHQSVQLPRE--GQ 420

Qy      322 EDGSCSYKDFSESNNRNFSTPEQAANK 348
      || : ||: | :|| | :||
Db      421 EDQGLT-KDYGSSPLHRFKKP--GSKN 444

```

# RESULT 7

S68857

polypyrimidine tract-binding protein - pig

C;Species: Sus scrofa domestica (domestic pig)

C;Date: 15-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 09-Jul-2004

C;Accession: S68857

R;Niepmann, M.

FEBS Lett. 388, 39-42, 1996

A;Title: Porcine polypyrimidine tract-binding protein stimulates translation initiation at the internal ribosome entry site of foot-and-mouth-disease virus.

A;Reference number: S68857; MUID:96249475; PMID:8654585

A;Accession: S68857

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-557 <NIE>

A;Cross-references: UNIPROT:Q29099; EMBL:X93009; NID:g1122432; PIDN:CAA63597.1;

PID:e213436; PID:g1122433

Query Match 18.0%; Score 345; DB 2; Length 557;

Best Local Similarity 28.0%; Pred. No. 3.4e-19;

Matches 111; Conservative 62; Mismatches 132; Indels 92; Gaps 14;

```

Qy      8 VNYYAADNQIYIAGHPAFVNYSTSQKISRPGSDDDSR-----SVNS----- 47
      ||| : | | :: :| :: : :| :|||
Db      111 VNYYTSVTPVLRGQPIYIQFSNHKELKTDSSPNQARAQAALQAVNSVQSGNLALAASAAA 170

Qy      48 -----VLLFTILNPIYSITTDVLYTICNPCGVPQRIVIFRKNQ-VQAMVEFDS 94
      || : | | :| |||: | : | | :| :| ||| :|:::
Db      171 VDAGMAMAGQSPVLRRIIVENLFYPVTLVDVLHQIFSFGTVLKIITFTKNNQFQALLQYAD 230

Qy      95 VQSAQRAKASLNGADIYSGCCTLKIEYAKPTRLNVFKNDQDTWDYTNPNLSGQGDPGSNP 154
      ||| || ||:| :||: ||||:|:::| | ||| |: : ||| |:| ||
Db      231 PVSAQHAKLSLDGQNIYNACCTLRIDFSKLTSLNVKYNNDKSRDYTRPDLP-SGD----- 284

Qy      155 NKRQRQPPLLGDHPAEYGGPHGGYHSHYHDEGYGPPP-----PHYEGRRMGP- 201
      || | | :| | | | | | | | | | :| : | :|
Db      285 ----NQPSLDQTMAAAFAGAP--GIMSASPYAGAGFPPTFAIPQAATVSVPNVHG-ALAPL 337

Qy      202 -----PVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMVYGLDQSKMNCDR 252
      | | | | | | | | | | | : ||:| |: ::
Db      338 AIPSAARAAAAGRIAPGLAG-----AGNSVLLVSNLNPERVTPQS 379

Qy      253 VFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFGQKLNVCVSK-QP 311
      :| :| :| :|::|| : :| |::||| |::||| : : |: : : :|| |

```



Db 380 LFILFGVYCDVQVRVKILFNKKENALVQMADGSQAQLAMSHLNGHKLHGKPVRLITLSKHQN 439

Qy 312 AIMGQSYGLEDGSCSYKDFSESNNRFSTPEQAANK 348  
 : | : | | : | | : | : | : | : |

Db 440 VQLPRE--GQEDQGLT-KDYGNSPLHRFKKP--GSKN 471

# RESULT 8

JC7526

polypyrimidine tract-binding protein-like protein - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 07-Jul-2003

C;Accession: JC7526

R;Kikuchi, T.; Ichikawa, M.; Arai, J.; Tateiwa, H.; Fu, L.; Higuchi, K.; Yoshimura, N.

J. Biochem. 128, 811-821, 2000

A;Title: Molecular cloning and characterization of a new neuron-specific homologue of rat polypyrimidine tract binding protein.

A;Reference number: JC7526; MUID:20512059; PMID:11056394

A;Contents: Neonatal retina

A;Accession: JC7526

A;Molecule type: mRNA

A;Residues: 1-532 <KIK>

A;Cross-references: GB:AJ010585

C;Comment: This protein is a retinal and neuron-specific protein that plays an important role in the development and alternative splicing in the neuronal cells. It also has multiple functions in the cytoplasm and nucleus during neurogenesis.

C;Genetics:

A;Gene: ptb1p

Query Match 17.9%; Score 343.5; DB 2; Length 532;

Best Local Similarity 26.1%; Pred. No. 4.2e-19;

Matches 102; Conservative 65; Mismatches 127; Indels 97; Gaps 11;

Qy 4 ACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRSV----- 45  
 | ||| : : : | : : | : : : : |

Db 107 AITMVNYYSAVTPHLRNQPIYIQYSNHKELKTDNTLNQRAQVVLQAVTAVQTANTPLSGT 166

Qy 46 -----NSVLLFTILNPIYSITTDVLYTICNPGPVQRIVIFRKNG-VQAMVEFD 93  
 : || | | | : | ||| : | : | | : | | | ||| : :

Db 167 TVSESAVTPAQSPVLRIIIDNMYYPVTLDLVLHQIFSKFGAVLKIITFTKNNQFQALLQYG 226

Qy 94 SVQSAQRAKASLNGADIYSGCCTLKIEYAKPTRLNVFKNDQDTWDYTNPNL-SGQGDPGS 152  
 : || : || : || : ||| : ||| : ||| : ||| : ||| : ||| : |||

Db 227 DPVNAQQAKLALDQNIYNACCTLRIDFSKLVNLNVKYNNDKSRDYTRPDLPSGDGQPAL 286

Qy 153 NPN-----KRQRQPPLLGDHPEYGGPHGGYHSHYHDEGYGPPPPHYEGRR 198  
 : | : : | : : |

Db 287 DPAIAAAFAKETSL LAVPGALSPLAIPNAAAAAAAAAAG-----R 326

Qy 199 MGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMVYGLDQSKMNCDRVFNVFC 258  
 : | | | | : : : : : | : | : | : | : |

Db 327 VGMP-----GVSAGG-----NTVLLVSNLNEEMVTPQSLFTLFG 360

Qy 259 LYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFGQKLNVCVSKQPAI-MPGQ 317  
 : || : || : || : ||| : ||| : ||| : ||| : ||| : ||| : |||



```

      | | | : ||:| | : : : | : | : ||:| : : |
Db      326 -----IPGLAG--AGNSVLLVSNLNPERVTPQSLFILFGVYGDVQRVKILFNKKE 373

Qy      275 AAMVEMADGYAVDRAITHLNNNFMFQGKLNVCVSKQPAI-MPGQSYGLEDGSCSYKDFSE 333
      |:|:| || | : : ||:| : : : | : : | : | | : ||:
Db      374 NALVQMADGSQAQLAMSHLNGHKLHGKSVRITLSKHQSVQLPRE--GQEDQGLT-KDYGS 430

Qy      334 SRNNRFSTPEQAAKN 348
      | | | : ||
Db      431 S-PLRFKKP--GSKN 442

```

# RESULT 10

A88299

protein D2089.4 [imported] - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004

C;Accession: A88299

R;anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology.

A;Reference number: A75000; MUID:99069613; PMID:9851916

A;Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www\_sanger.ac.uk/Projects/C\_elegans/ for a list of authors

A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and Science 285, 1493, 1999

A;Accession: A88299

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-584 <STO>

A;Cross-references: UNIPROT:Q18999; GB:chr\_II; PIDN:CAA85411.1; PID:g3875368; GSPDB:GN00020; CESP:D2089.4

A;Note: similar to polypyrimidine tract binding protein

C;Genetics:

A;Gene: D2089.4

A;Map position: 2

Query Match 15.4%; Score 296.5; DB 2; Length 584;

Best Local Similarity 26.0%; Pred. No. 2.1e-15;

Matches 100; Conservative 57; Mismatches 124; Indels 103; Gaps 14;

```

Qy      19 AGHPAFVNYSTSQKISRPGSDDDS-----RSVNSVLLFTILNPIYSITTDVLYTICNP 71
      | ||: | : | : | : ||| | | : : ||| :
Db      164 ASAAAFVSGMTAVPIQSVANGSVSNFEVGTQQQPNSVLRTIIENMMFPVSLDVLYQLFTR 223

Qy      72 CGPVQRIVIFRKNQV-QAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIEYAKPTRLNVF 130
      | | ||: | || ||:| : ||| || | :||:|||||:|:| : |||
Db      224 YGKVLRIITFNKNNTFQALVQMSEANSAQLAKQGLENQNVYNGCCTLRIDYSKLSTLNVK 283

Qy      131 KNDQDTWDYTNPNL-SGQ-----GDP-----G SNP----- 154
      | : : ||||| : | : | : |||
Db      284 YNNDKSRDYTNPNLPAGEMTLEQTIAMSIPGLQNLI PANPYNFAFGANPATTFLTTLQALAA 343

Qy      155 -----NKRQRQPPLLDGHPAEYGGPHGGYHSHYHDEGYGPPPPHYEGRMGPPVGGH 206
      | | | | | | : | :
Db      344 STAAAAAVNDSANAAAL-----APYLNPLG-----LTSANLAPSISSM 381

```

```

Qy      207 RRGPSRYGPQYGHPPPPPPPEYGPHAD-SPVLMVYGLDQSKMNCDRVFNVFCLYGNVEK 265
      |      |      |      |      |      |      |      |      |      |      |      |      |
Db      382 R-----FPMINLTPVILVSNLHEMKVTTDALFTLFGVYGDVMR 419

Qy      266 VKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMSGQKLNVCVSKQPAI-MPGQSYGLEDG 324
      || : :|  |::: ::  |::||:  : : |  ||  : || : | |
Db      420 VKILYNKKDNALIQYSEPQQAQLALTHLDKVKWHDRLIRVAPSKHTNVQMPKE--GQPD 477

Qy      325 SCSYKDFSESRNNRFSTPEQAAKN 348
      : :||: |  :||  |  :||
Db      478 GLT-RDYAHSTLHRFKKP--GSKN 498

```

# RESULT 11

T20381

hypothetical protein D2089.4 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C;Accession: T20381

R;Swinburne, J.

submitted to the EMBL Data Library, September 1994

A;Reference number: Z19264

A;Accession: T20381

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-592 <WIL>

A;Cross-references: UNIPROT:Q18999; EMBL:Z36948; PIDN:CAA85411.2; GSPDB:GN00020;  
CESP:D2089.4

A;Experimental source: clone D2089

C;Genetics:

A;Gene: CESP:D2089.4

A;Map position: 2

A;Introns: 3/3; 98/3; 126/3; 163/3; 187/3; 245/1; 319/1; 361/1; 408/3; 420/1;  
451/1; 549/2

Query Match 15.4%; Score 296.5; DB 2; Length 592;

Best Local Similarity 26.0%; Pred. No. 2.1e-15;

Matches 100; Conservative 57; Mismatches 124; Indels 103; Gaps 14;

```

Qy      19 AGHPAFVNYSTSQKISRPGSDSDS-----RSVNSVLLFTILNPIYSITTDVLYTICNP 71
      |  |||: | : |  :  |  :  |||  | | : : ||| :
Db      172 ASAAAFVSGMTAVPIQSVANGSVSNFEVGTQQQPNSVLRTIENMMFPVSLDVLYQLFTR 231

Qy      72 CGPVQRIVIFRKNGV-QAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIEYAKPTRLNVF 130
      | | ||: | ||  ||::  ||| ||  |  ::||| |||:|:| : |||
Db      232 YGKVLRIITFNKNNTFQALVQMSEANSAQLAKQGLENQNVYNGCCTLRIDYSKLSTLNVK 291

Qy      131 KNDQDTWDYTNPNL-SGQ-----GDP-----GSPN----- 154
      | :  : ||||| :|:  :|  |::|
Db      292 YNNDKSRDYTNPNLPAGEMTLEQTIAMSIPGLQNLIPANPYNFAFGANPATTFLTTLQAA 351

Qy      155 -----NKRQRQPPLLDGHPAEYGGPHGGYHSHYHDEGYGPPPPHYEGRRMGPPVGGH 206
      |      |  |  |  |  |  :  | :
Db      352 STAAAAAVNDSANAAAL-----APYLNPLG-----LTSANLAPSISSM 389

Qy      207 RRGPSRYGPQYGHPPPPPPPEYGPHAD-SPVLMVYGLDQSKMNCDRVFNVFCLYGNVEK 265

```

```

      |           | : :|:|:| | : | : | :| :| :
Db      390 R-----FPMINLTPVILVSNLHEMKVTTDALFTLFGVYGDVMR 427

Qy      266 VKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFQGKLNVCVSKQPAI-MPGQSYGLEDG 324
      || : :| |::: :: |::||: : : | || : || : | |
Db      428 VKILYNKKDNALIQYSEPQQAQLALTHLDKVKWHDRLIRVAPSKHTNVQMPKE--GQPD 485

Qy      325 SCSYKDFSESRRNRFSTPEQAAKN 348
      : :|:|:| | :|| | :||
Db      486 GLT-RDYAHSTLHRFKKP--GSKN 506

```

# RESULT 12

T51814

polypyrimidine tract-binding protein homolog [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 09-Jul-2004

C;Accession: T51814

R;Marin, C.; Boronat, A.

submitted to the EMBL Data Library, July 1998

A;Reference number: Z25464

A;Accession: T51814

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-418 <MAR>

A;Cross-references: UNIPROT:O82472; EMBL:AF076924; PIDN:AAC62015.1

C;Genetics:

A;Gene: PTB

Query Match 11.3%; Score 217.5; DB 2; Length 418;

Best Local Similarity 33.2%; Pred. No. 2e-09;

Matches 65; Conservative 32; Mismatches 76; Indels 23; Gaps 8;

```

Qy      2 LGACN-AVNYYA--DNQIYIAGHPAF-----VNYSTSQKISRPGDSDDSRSVNS--- 47
      :|:|:| :|:|:| | | | : | | : | : | |
Db      184 VGSCSLRMSYSAHTDLNLIKFSHRSDYTNPYLPVNQTAMDGSMQPALGADGKKVESQSN 243

Qy      48 VLLFTILNPIYSITTDVLYTICNPGCPVQRIVIFRKNQ-VQAMVEFDSVQSAQRAKASLN 106
      ||| | | |::| ||:|: : | ||:| || || |::: : :| || :|
Db      244 VLLGLIENMQYAVTVDLHTVFSAYGTQVKIAIFEKNGSTQALIQYSDIPTAAMAKEALE 303

Qy      107 GADIY-SGCCTLKIEYAKPTRLNVFKNDQDTWDYTNPNLS-----GQGDPGSNPNKRQR 159
      | || | | :|:|:| | || : ||| |:| | | | | :
Db      304 GHCIYDGGYCKLRLSYSRHTDLNVKAFSDKSRDYTLPLDLLVAQKGPVSGSAPPAGWQ 363

Qy      160 QPPLLGDHPAEYGGPH 175
      | : : ||| |
Db      364 NPQAQSQY-SGYGGSH 378

```

# RESULT 13

T10015

hypothetical protein MLB1770.15c - Mycobacterium leprae

C;Species: Mycobacterium leprae

C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 09-Jul-2004

C;Accession: T10015

R;Cole, S.T.

submitted to the EMBL Data Library, August 1997

A;Reference number: Z16916

A;Accession: T10015

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-463 <COL>

A;Cross-references: UNIPROT:Q50190; EMBL:Z70722; NID:e1059634; PID:e337961

C;Genetics:

A;Note: MLB1770.15c

```
Query Match          8.0%; Score 154; DB 2; Length 463;
Best Local Similarity 24.8%; Pred. No. 0.0002;
Matches 69; Conservative 20; Mismatches 105; Indels 84; Gaps 13;

Qy      139 YTNPNLSGQG-DPGSNPNKRQRPPLLGDHPAEYGGP-----HGGYHSH--YH 183
      | | :| | | : | | | :| : | | | | | | | | | |
Db      151 YGRPQDDPRGADPQGGQDPRGCYPPKPGSYQQAGHPPLHRPDQGGYPGQGGYEDQRAYH 210

Qy      184 DEGYGPPPPHYEGR-----RMGPPVGG-----HRRGPSR--- 212
      |:| | | | | | | | | | | | | | | | | | | |
Db      211 DQGQGGYPSPYEQRPATPGGYGSQGHQGYRPGSYGPPSGGQPGYGGYGDYGRGPARPDE 270

Qy      213 --YGPQYGHPPPPPPP---PEYGPHADSPVLMVYGLDQSKMNCDRVFNVFCLYGNVEKVK 267
      | | | | | | | | | | | | | | | | | | | |
Db      271 GSYTPS-GFPAPPEQRVAYPDQGGGYDQ-----GYQHSGLGYGRED-----YGRQEYTO 318

Qy      268 FMKSKPGAAMVEMADGYAVDRAITHLNNNFMTFGQKLNVCVSKQPAIMPGQSYG----- 320
      : :: || : || | : : : || | | | | | |
Db      319 YAENLPGGVYAPSSGGYA-----EPAGRDYDYGQPGAANDYSQPVIGGYGGYGALGSAVI 373

Qy      321 --LEDGSCSYKDFSES RN-----NRFSTPEQA AKNR 349
      |:| | | | | | | | | | | | | | | |
Db      374 LQLDDGSGRTYQLREGSNIVGRGQDAQFRLPDTGVSRR 411
```

#### RESULT 14

F86911

conserved hypothetical protein ML0022 [imported] - Mycobacterium leprae

C;Species: Mycobacterium leprae

C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004

C;Accession: F86911

R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Honore, N.; Ganier, T.; Churcher, C.; Harris, D.; Mungall, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Lacroix, C.; Maclean, J.; Moule, S.; Murphy, L.; Oliver, K.; Quail, M.A.; Rajandream, M.A.; Rutherford, K.M.

Nature 409, 1007-1011, 2001

A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Squares, S.; Stevens, K.; Taylor, K.; Whitehead, S.; Woodward, J.R.; Barrell, B.G.

A;Title: Massive gene decay in the leprosy bacillus.

A;Reference number: A86909; MUID:21128732; PMID:11234002

A;Accession: F86911

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-488 <STO>

A;Cross-references: UNIPROT:Q9CDE4; GB:AL450380; NID:g13092432; PIDN:CAC29530.1;  
GSPDB:GN00147  
C;Genetics:  
A;Gene: ML0022

Query Match 8.0%; Score 154; DB 2; Length 488;  
Best Local Similarity 24.8%; Pred. No. 0.00021;  
Matches 69; Conservative 20; Mismatches 105; Indels 84; Gaps 13;

```
Qy      139 YTNPNLSGQG-DPGSNPNKRQRPPLLGDHPAEYGGP-----HGGYHSH--YH 183
      | | :| | | : | | | :| : | | | | |
Db      176 YGRPQDDPRGADPQGGQDPRGCPYPPKPGSYPPQAGHPPLHRPDQGGYPGQGGYEDQRAYH 235

Qy      184 DEGYGPPPPHYEGR-----RMGPPVGG-----HRRGPSR--- 212
      |:| | | | | | | | | | | | | | | | |
Db      236 DQGQGGYPSPYEQRPATPGGYGSQGHQGYRPGSYGPPSGGQPGYGGYGDYGRGPARPDE 295

Qy      213 --YGPQYGHPPPPPPP---PEYGPHADSPVLMVYGLDQSKMNCDRVFNVFCLYGNVEKVK 267
      | | | | | | | | | | | | | | | | |
Db      296 GSYTPS-GFPAPPEQRVAYPDQGGGYDQ-----GYQHSGLGYGRED-----YGRQEYDQ 343

Qy      268 FMKSKPGAAMVEMADGYAVDRAITHLNNNFMFQGLNVCVSKQPAIMPGQSYG----- 320
      : :: | | : | | | : : | | | | | |
Db      344 YAENLPGGVYAPSSGGYA-----EPAGRDYDYGQPGAANDYSQPVIGGYGGYGALGSAVI 398

Qy      321 --LEDGSCSYKDFSES RN-----NRFSTPEQA AKNR 349
      |:| | | | | | | | | | | | |
Db      399 LQLDDGSGRTYQLREGSNIVGRGQDAQFRLPDTGVSRR 436
```

#### RESULT 15

T15264

hypothetical protein F59E12.9 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004

C;Accession: T15264

R;Johnson, D.

submitted to the EMBL Data Library, May 1997

A;Description: The sequence of *C. elegans* cosmid F59E12.

A;Reference number: Z18318

A;Accession: T15264

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-1621 <JOH>

A;Cross-references: UNIPROT:O01900; EMBL:AF003386; NID:g2088833; PID:g2088843;

PIDN:AAB54259.1; GSPDB:GN00020; CESP:F59E12.9

A;Experimental source: strain Bristol N2; clone F59E12

C;Genetics:

A;Gene: CESP:F59E12.9

A;Map position: 2

A;Introns: 30/3; 55/1; 200/2; 299/2; 327/2; 369/3; 589/3; 860/1; 986/1; 1278/1;  
1547/1

Query Match 7.9%; Score 152.5; DB 2; Length 1621;  
Best Local Similarity 37.6%; Pred. No. 0.0011;  
Matches 41; Conservative 7; Mismatches 42; Indels 19; Gaps 6;

Qy 142 PNLSGQGDPSN---PNKRQRPPLLGDHPAEYGGPHGGYHSHYHD--EGYGP-----PP 191  
 | : | | | : | : | | | | : | | : | | |  
 Db 1502 PPMFRGGPPGPGRGMPSPMMRGSSMRGGFPQRGGGPGMGPSQYYHDSPQNRGPPMGGLPP 1561

Qy 192 PH--YEGRRMGPPV---GGHRRGPSRY----GPQYGHPPPPPPPEYGP 231  
 | | | | | | : | | : | | | | | | |  
 Db 1562 PHGGMNGWRGGPPPPRGGSHCQGPPLMGGPPPRLGMPPPGPPPPNGGP 1610

Search completed: January 7, 2005, 14:52:24  
 Job time : 20.8061 secs



OM protein - protein search, using sw model

Run on: January 7, 2005, 14:51:07 ; Search time 65.6283 Seconds  
(without alignments)  
1917.457 Million cell updates/sec

Title: US-10-726-721A-7  
Perfect score: 1921  
Sequence: 1 VLGACNAVNYAADNQIYIAG.....DFSESRRNRFSTPEQAAKNR 349

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1603904 seqs, 360571292 residues

Total number of hits satisfying chosen parameters: 1603904

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep:\*  
19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1921	100.0	349	9	US-09-780-996-7	Sequence 7, Appli
2	1921	100.0	349	16	US-10-726-721-7	Sequence 7, Appli
3	1909	99.4	589	14	US-10-353-929-46	Sequence 46, Appl
4	780	40.6	301	9	US-09-925-301-1354	Sequence 1354, Ap
5	426	22.2	168	15	US-10-108-260A-4694	Sequence 4694, Ap
6	390	20.3	444	17	US-10-425-115-199139	Sequence 199139,
7	390	20.3	481	15	US-10-425-114-60710	Sequence 60710, A
8	389	20.2	444	17	US-10-425-115-199137	Sequence 199137,
9	389	20.2	481	15	US-10-425-114-62527	Sequence 62527, A
10	349.5	18.2	550	9	US-09-895-828-452	Sequence 452, App
11	349.5	18.2	550	14	US-10-114-666-452	Sequence 452, App
12	343.5	17.9	532	14	US-10-205-219-163	Sequence 163, App
13	329.5	17.2	521	16	US-10-322-281-292	Sequence 292, App
14	324	16.9	322	16	US-10-408-765A-1921	Sequence 1921, Ap
15	312	16.2	297	16	US-10-437-963-199300	Sequence 199300,
16	302.5	15.7	482	16	US-10-322-281-289	Sequence 289, App
17	283	14.7	316	15	US-10-424-599-268658	Sequence 268658,
18	257	13.4	230	15	US-10-424-599-270019	Sequence 270019,
19	254.5	13.2	487	17	US-10-739-930-8044	Sequence 8044, Ap
20	253	13.2	239	17	US-10-425-115-198637	Sequence 198637,
21	249	13.0	786	16	US-10-437-963-134250	Sequence 134250,
22	236.5	12.3	375	9	US-09-925-300-1674	Sequence 1674, Ap
23	236	12.3	239	17	US-10-425-115-199136	Sequence 199136,
24	232.5	12.1	283	15	US-10-424-599-272023	Sequence 272023,
25	231.5	12.1	548	16	US-10-437-963-199298	Sequence 199298,
26	226.5	11.8	554	16	US-10-437-963-170013	Sequence 170013,
27	220.5	11.5	298	15	US-10-424-599-176841	Sequence 176841,
28	220	11.5	433	17	US-10-425-115-336341	Sequence 336341,
29	220	11.5	434	15	US-10-425-114-57992	Sequence 57992, A
30	220	11.5	434	15	US-10-425-114-70964	Sequence 70964, A
31	214.5	11.2	465	17	US-10-425-115-274072	Sequence 274072,
32	213	11.1	333	15	US-10-425-114-57954	Sequence 57954, A
33	213	11.1	488	15	US-10-425-114-60130	Sequence 60130, A
34	213	11.1	611	17	US-10-425-115-274071	Sequence 274071,
35	212	11.0	230	16	US-10-437-963-188788	Sequence 188788,
36	202	10.5	429	15	US-10-424-599-176833	Sequence 176833,
37	192	10.0	408	17	US-10-425-115-317531	Sequence 317531,
38	190.5	9.9	340	15	US-10-424-599-176839	Sequence 176839,
39	189	9.8	466	17	US-10-425-115-274063	Sequence 274063,
40	171	8.9	213	17	US-10-739-930-8998	Sequence 8998, Ap
41	169.5	8.8	219	16	US-10-437-963-134244	Sequence 134244,
42	167.5	8.7	122	16	US-10-767-701-40635	Sequence 40635, A
43	166	8.6	349	15	US-10-424-599-166513	Sequence 166513,
44	157	8.2	25	14	US-10-197-857-2	Sequence 2, Appli
45	157	8.2	371	17	US-10-425-115-320892	Sequence 320892,

#### ALIGNMENTS

##### RESULT 1

US-09-780-996-7

; Sequence 7, Application US/09780996

; Patent No. US20020061553A1

```
; GENERAL INFORMATION:
; APPLICANT: Maury, Isabella
; APPLICANT: Mercken, Luc
; APPLICANT: Fournier, Alain
; TITLE OF INVENTION: Partners of the PTB1 Domain of FE65, Preparation and Uses
; FILE REFERENCE: ST00004-US
; CURRENT APPLICATION NUMBER: US/09/780,996
; CURRENT FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: FR00/01628
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: US 60/198,500
; PRIOR FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-780-996-7
```

```
Query Match          100.0%; Score 1921; DB 9; Length 349;
Best Local Similarity 100.0%; Pred. No. 9.3e-160;
Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRSVNSVLLFTILNPIYSI 60
        |||
Db      1 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRSVNSVLLFTILNPIYSI 60

Qy     61 TTDVLYTICNPCGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 120
        |||
Db     61 TTDVLYTICNPCGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 120

Qy    121 YAKPTRLNVFKNDQDQTDWDTNPNLSGQGDPSNPKNRQRPPLLGDHPAEYGGPHGGYHS 180
        |||
Db    121 YAKPTRLNVFKNDQDQTDWDTNPNLSGQGDPSNPKNRQRPPLLGDHPAEYGGPHGGYHS 180

Qy    181 HYHDEGYGPPPPHYEGRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV 240
        |||
Db    181 HYHDEGYGPPPPHYEGRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV 240

Qy    241 YGLDQSKMNCNDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFG 300
        |||
Db    241 YGLDQSKMNCNDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFG 300

Qy    301 QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESNNRNFSTPEQAAKNR 349
        |||
Db    301 QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESNNRNFSTPEQAAKNR 349
```

# RESULT 2

US-10-726-721-7

```
; Sequence 7, Application US/10726721
; Publication No. US20040166109A1
; GENERAL INFORMATION:
; APPLICANT: Maury, Isabella
; APPLICANT: Mercken, Luc
; APPLICANT: Fournier, Alain
```

```

; TITLE OF INVENTION: Partners of the PTB1 Domain of FE65, Preparation and Uses
; FILE REFERENCE: ST00004-US
; CURRENT APPLICATION NUMBER: US/10/726,721
; CURRENT FILING DATE: 2003-12-03
; PRIOR APPLICATION NUMBER: US/09/780,996A
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: FR00/01628
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: US 60/198,500
; PRIOR FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-726-721-7

```

```

Query Match          100.0%; Score 1921; DB 16; Length 349;
Best Local Similarity 100.0%; Pred. No. 9.3e-160;
Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSI 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSI 60

Qy     61 TTDVLYTICNPCGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TTDVLYTICNPCGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 120

Qy    121 YAKPTRLNVFKNDQDTWDYTNPNLSGQDPSNPKNRQRPPLLGDHPAEYGGPHGGYHS 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 YAKPTRLNVFKNDQDTWDYTNPNLSGQDPSNPKNRQRPPLLGDHPAEYGGPHGGYHS 180

Qy    181 HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV 240

Qy    241 YGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFG 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 YGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFG 300

Qy    301 QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESNNRNFSTPEQAAKNR 349
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESNNRNFSTPEQAAKNR 349

```

```

RESULT 3
US-10-353-929-46
; Sequence 46, Application US/10353929
; Publication No. US20030175288A1
; GENERAL INFORMATION:
; APPLICANT: ITOH, Kyogo
; TITLE OF INVENTION: Tumor antigen
; FILE REFERENCE: GP01-1024
; CURRENT APPLICATION NUMBER: US/10/353,929
; CURRENT FILING DATE: 2003-01-30

```

; PRIOR APPLICATION NUMBER: JP P2000-231814  
; PRIOR FILING DATE: 2000-07-31  
; NUMBER OF SEQ ID NOS: 197  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 46  
; LENGTH: 589  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-353-929-46

Query Match 99.4%; Score 1909; DB 14; Length 589;  
Best Local Similarity 99.7%; Pred. No. 2e-158;  
Matches 348; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSI 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     147 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSI 206

Qy      61 TTDVLYTICNPCGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     207 TTDVLYTICNPCGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 266

Qy     121 YAKPTRLNVFKNDQDTWDYTNPNLSGQDGPSPNPNKRQRPPLLGDHPAEYGGPHGGYHS 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     267 YAKPTRLNVFKNDQDTWDYTNPNLSGQDGPSPNPNKRQRPPLLGDHPAEYGGPHGGYHS 326

Qy     181 HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     327 HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV 386

Qy     241 YGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFG 300
        |||||||| ||||||||||||||||||||||||||||||||||||||||
Db     387 YGLDQSKMNGDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFG 446

Qy     301 QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESNNRNFSTPEQAAKNR 349
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db     447 QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESNNRNFSTPEQAAKNR 495
```

RESULT 4

US-09-925-301-1354

; Sequence 1354, Application US/09925301

; Patent No. US20020052308A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

; FILE REFERENCE: PA106

; CURRENT APPLICATION NUMBER: US/09/925,301

; CURRENT FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: PCT/US00/05882

; PRIOR FILING DATE: 2000-03-08

; PRIOR APPLICATION NUMBER: 60/124,270

; PRIOR FILING DATE: 1999-03-12

; NUMBER OF SEQ ID NOS: 1694

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1354

; LENGTH: 301

; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-925-301-1354

Query Match 40.6%; Score 780; DB 9; Length 301;  
Best Local Similarity 99.3%; Pred. No. 7.3e-60;  
Matches 148; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDDDSRVNSVLLFTILNPIYSI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      137 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDDDSRVNSVLLFTILNPIYSI 196

Qy      61 TTDVLYTICNPCGPVQRIVIFRKNVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      197 TTDVLYTICNPCGPVQRIVIFRKNVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 256

Qy      121 YAKPTRLNVFKNDQDTWDYTNPNLSGQGD 149
          |||||||||||||||||||:
Db      257 YAKPTRLNVFKNDQDTWDYTNPNLSGQGN 285
```

RESULT 5

US-10-108-260A-4694  
; Sequence 4694, Application US/10108260A  
; Publication No. US20040005560A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. US20040005560A1el full length cDNA  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/10/108,260A  
; CURRENT FILING DATE: 2002-03-27  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4694  
; LENGTH: 168  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-108-260A-4694

Query Match 22.2%; Score 426; DB 15; Length 168;  
Best Local Similarity 100.0%; Pred. No. 3.6e-29;  
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      269 MKSKPGAAMVEMADGYAVDRAITHLNNNFMFGQKLNVCVSKQPAIMPGQSYGLEDGSCSY 328
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MKSKPGAAMVEMADGYAVDRAITHLNNNFMFGQKLNVCVSKQPAIMPGQSYGLEDGSCSY 60

Qy      329 KDFSESRNNRFSTPEQAAKNR 349
          ||||||||||||
Db      61 KDFSESRNNRFSTPEQAAKNR 81
```

RESULT 6

US-10-425-115-199139  
; Sequence 199139, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:

```

; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated
With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 199139
; LENGTH: 444
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_113190C.1.pep
US-10-425-115-199139

```

```

Query Match          20.3%; Score 390; DB 17; Length 444;
Best Local Similarity 28.9%; Pred. No. 1.8e-25;
Matches 101; Conservative 65; Mismatches 118; Indels 66; Gaps 7;

```

```

Qy      6 NAVNYAADNQIYIAGHPAFVNYSTSQKI-----SRPGDSDDSRVNSVLLFTILNPIYSI 60
      :|: |   |   |   :: :|: |:|   |   :||   | :|| || : || |
Db      56 SALQYYTSVQPSIRGRNVYMQFSSHQELTTDQSSHGRNSDQGSEPNRILLVTIHHMIYPI 115

Qy      61 TTDVLYTICNPCGPVQRIVIFRKN-GVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKI 119
      | ::|: :   | |::|| |:|: | ||::: | | | | ||:| :|| || | |
Db      116 TVEILHQVFKAIGFVEKIVTFQKSAGFQALIQYHSRQEAVEAFGSLHGRNIYDGCCQLDI 175

Qy      120 EYAKPTRLNVFKNDQDTWDYTNPNLSGQDGPSPNKRQRQPPLLGDHPAEYGGPHGGYH 179
      :|: : | | | : : |:| ||:| : | ::
Db      176 QYSNLSELQVHYNNDRSRDFTNPSLPTEQRPAS----- 209

Qy      180 SHYHDEGYGPPPPHYEGRRMG---PPVGGHRRGPSRYGPQYGHPPPPPPPEYGPFA--- 233
      :|| |   |   :: |   :|   : :|   | |
Db      210 ----QQGYLDPANLYAFQQAGASYAQMGVRVAMIAAAFGGTL-----PHGVTG 252

Qy      234 --DSPVLMVYGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAIT 291
      :   |:| |: |:|: |:|:| | |||: ::| ::|| |:| |||| : |:
Db      253 TNERCTLIVSNLNTDKIDEDKLFNLSLYGNIVRIKILRNKPDHALVEMADGLQAE LAVH 312

Qy      292 HLNNFMFGQKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESRRNRFST 341
      :| : :||:| | | | | | | | : | | |::
Db      313 YLKGSI LFGKKLEVNYSKYPNITPAP-----DAHDYLNSSINRFNS 353

```

```

RESULT 7
US-10-425-114-60710
; Sequence 60710, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E

```

```
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated
With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 60710
; LENGTH: 481
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3587-267-C11_FLI.pep
US-10-425-114-60710
```

```
Query Match          20.3%; Score 390; DB 15; Length 481;
Best Local Similarity 28.9%; Pred. No. 2e-25;
Matches 101; Conservative 65; Mismatches 118; Indels 66; Gaps 7;
```

```
Qy      6 NAVNYAADNQIYIAGHPAFVNYSTSQKI-----SRPGDSDDSRVNSVLLFTILNPIYSI 60
      :|: |   |   |   : : :|: |:   |   :|   |   |   |   |
Db      93 SALQYYTSVQPSIRGRNVYMQFSSHQELTTDQSSHGRNSDQGSEPNRILLVTIHMIYPI 152

Qy      61 TTDVLYTICNPCGPVQRIVIFRKN-GVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKI 119
      | : :|: :   | | : :| | : :| | : : :| | |   |   |   |
Db      153 TVEILHQVFKA YGFVEKIVTFQKSAGFQALIQYHSRQEAVEAFGSLHGRNIYDGCCQLDI 212

Qy      120 EYAKPTRLN VFKNQDQDTWDYTNPNLSGQDPGSNPNKRQRQPPLLGDHPAEYGGPHGGYH 179
      :|: : | | | : : :|: |||: | : | : :
Db      213 QYSNLSELQVHYNNDRSRDFTNPSLPTEQRPAS----- 246

Qy      180 SHYHDEGYGPPPPHYEGRRMG---PPVGGHRRGPSRYGPQYGHPPPPPPPEYGPFA--- 233
      :| | |   |   : : |   :|   : :|   |
Db      247 ----QQGYLDPANLYAFQQAGASYAQMG RVAMIAAAFGGTL-----PHGVTG 289

Qy      234 --DSPVLMVYGLDQSKMNC DRVFN VFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAIT 291
      :   | :| | : | : :|: |||: :| : :|| | :| |||| | : | :
Db      290 TNERCTLIVSNLNTDKIDEDKLFNLFSLYGNIVRIKILRNKPDHALVEMADGLQAE LAVH 349

Qy      292 HLNNNFMFQKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESRRNNRFST 341
      :| : :|: || | | | | |   | : | | | : :
Db      350 YLKGSI LFGKKLEVNYSKYPNITPAP-----DAHDYLNSSINRFNS 390
```

# RESULT 8

US-10-425-115-199137

; Sequence 199137, Application US/10425115

; Publication No. US20040214272A1

## GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated  
With



; TITLE OF INVENTION: Plants  
 ; FILE REFERENCE: 38-21(53222)B  
 ; CURRENT APPLICATION NUMBER: US/10/425,115  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 369326  
 ; SEQ ID NO 199137  
 ; LENGTH: 444  
 ; TYPE: PRT  
 ; ORGANISM: Zea mays  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: MRT4577\_113189C.1.pep  
 US-10-425-115-199137

Query Match 20.2%; Score 389; DB 17; Length 444;  
 Best Local Similarity 29.1%; Pred. No. 2.2e-25;  
 Matches 102; Conservative 61; Mismatches 121; Indels 66; Gaps 7;

```

Qy      6 NAVNYAADNQIYIAGHPAFVNYSTSQKI-----SRPGDSDDSRSVNSVLLFTILNPIYSI 60
      :|: |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      56 SALQYYTSVQPSIRGRNVYMQFSSHQELTTDQSSHGRNSDQSEPNRILLVTIHHMIYPI 115

Qy      61 TTDVLYTICNPGCPVQRIVIFRKN-GVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKI 119
      | :||: :  | |::|| |::: | |:::| | | | | | | | | | | | | | |
Db      116 TVEVLHQVFKA YGFVEKIVTFQKSAGFQALIQFHSRQEAVEAFGSLHGRNIYDGCCQLDI 175

Qy      120 EYAKPTRLNVFKNDQDTWDYTNPNLSGQGDPGSNPNKRQRQPPLLGDHPAEYGGPHGGYH 179
      :|: : | | | : : |::||:| : | ::
Db      176 QYSNLSELQVHYNNDRSRDFTNPSLPTEQRPRAS----- 209

Qy      180 SHYHDEGYGPPPPHYEGRRMGPPVGGHRRG---PSRYGPQYGHPPPPPPPEYGPHA--- 233
      : | | | | : : |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      210 ----QQAYPD PANLYAFQQAGASYAQMGRAAMIAAAGGTL-----PHGVTG 252

Qy      234 --DSPVLMVYGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAIT 291
      :  |:| | : |:: |::||:| |::: :| :::|| |::||| | : |:
Db      253 TNERCTLIVSNLNNDKIDEDKLFNLFSLYGNIVRIKVLRNKPDHALVEMADGLQAE LAVH 312

Qy      292 HLNNNFMFGQKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESRRNRFST 341
      :|  :||:| | | | | | | | | | | | | | | | | | | | | |
Db      313 YLKGAILFGKKLEVNYSKYPNITPAP-----DAHDYLNSSLNRFNS 353
  
```

# RESULT 9

US-10-425-114-62527  
 ; Sequence 62527, Application US/10425114  
 ; Publication No. US20040034888A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Liu, Jingdong  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Screen, Steven E  
 ; APPLICANT: Tabaska, Jack E  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated  
 With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 62527  
; LENGTH: 481  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700470940\_FLI.pep  
US-10-425-114-62527

Query Match 20.2%; Score 389; DB 15; Length 481;  
Best Local Similarity 29.1%; Pred. No. 2.4e-25;  
Matches 102; Conservative 61; Mismatches 121; Indels 66; Gaps 7;

```
Qy      6 NAVNYAADNQIYIAGHPAFVNYSTSQKI-----SRPGDSDDSRVNSVLLFTILNPIYSI 60
      :|: | | | | :|: |:| :| | | | | | |
Db      93 SALQYYTSVQPSIRGRNVYMQFSSHQELTTDQSSHGRNSDQSEPNRILLVTIHHMIYPI 152

Qy      61 TTDVLYTICNPGPQVQIRIVIFRKN-GVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKI 119
      | :||: : | | :|| | :| | | | | | | | | | | |
Db      153 TVEVLHQVFKAYGFVEKIVTFQKSAGFQALIQFHSRQEAVEAFGSLHGRNIYDGCCQLDI 212

Qy      120 EYAKPTRLNVFKNDQDQDTWDYTNPNLSGQGDPGSNPNKRQRQPPLLGDHPAEYGGPHGGYH 179
      :|: : | | | : : | :|||:| : | :|
Db      213 QYSNLSELQVHYNNDRSRDFTNPSLPTEQRPAS----- 246

Qy      180 SHYHDEGYGPPPPHYEGRRMGPPVGGHRRG---PSRYGPQYGHPPPPPPPEYGPHA--- 233
      : | | | :| | :| :| :| |
Db      247 ----QQAYPDPANLYAFQQAGASYAQMGRAMIAAAFGGTL-----PHGVTG 289

Qy      234 --DSPVLMVYGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAIT 291
      : | :| | :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      290 TNERCTLIVSNLNNDKIDEDKLFNLFSLYGNIVRIKVLRNKPDHALVEMADGLQAE LAVH 349

Qy      292 HLNNNFMFGQKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESRRNRFST 341
      :| :| :| | | | | | | | | | | | | | | :| | | :|
Db      350 YLKGAILFGKKLEVNYSKYPNITPAP-----DAHDYLNSSLNRFNS 390
```

RESULT 10

US-09-895-828-452

; Sequence 452, Application US/09895828

; Patent No. US20020099012A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong

; APPLICANT: McNeill, Patricia D.

; APPLICANT: Watanabe, Yoshihiro

; APPLICANT: Carter, Darrick

; APPLICANT: Henderson, Robert A.

; APPLICANT: Kalos, Michael D.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER

; FILE REFERENCE: 210121.539

; CURRENT APPLICATION NUMBER: US/09/895,828

; CURRENT FILING DATE: 2001-06-28

; NUMBER OF SEQ ID NOS: 473

; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 452  
; LENGTH: 550  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-895-828-452

Query Match 18.2%; Score 349.5; DB 9; Length 550;  
Best Local Similarity 27.5%; Pred. No. 8.2e-22;  
Matches 109; Conservative 63; Mismatches 126; Indels 99; Gaps 14;

```
Qy      8 VNAAADNQIYIAGHPAFVNYSTSQKISRPGSDDDSR-----SVNS----- 47
      |||      : | | :: :|  ::      : :|      :|||
Db     111 VNYYTSVTPVLRGQPIYIQFSNHKELKTDSSPNQARAQALQAVNSVQSGNLALAASAAA 170

Qy     48 -----VLLFTILNPIYSITTDVLYTICNPCGPVQRIVIFRKNQ-VQAMVEFDS 94
      ||      : | | :| ||| : | : | | :| :| | ||| :| :| :|
Db     171 VDAGMAMAGQSPVLRRIIVENLFYPVTLVDVLHQIFSKFGTVLKIITFTKNNQFQALLQYAD 230

Qy     95 VQSAQRAKASLNGADIYSGCCTLKIEYAKPTRLNVFKNQDQTDWDTNPNLSGQGDPSGNSP 154
      ||| || ||| :| :| :| ||| :| :| :| ||| :| :| |||
Db     231 PVSAQHAKLSLDGQNIYNACCTLRIDFSKLTSLNVKYNNDKSRDYTRPDLP-SGD----- 284

Qy    155 NKRQRQPPLLGDHPAEYGGPHGGYHSHYHDEGYGPPP-----PHYEGRMGP- 201
      || |      | : | :      | |||      | : | : |
Db     285 ----SQPSLDQTMAAAFASPYA-----GAGFPPTFAIPQAAGLSVPNVHG-ALAPL 330

Qy    202 -----PVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMVYGLDQSKMNCDR 252
      |      |      |      | : ||| :| :| :|
Db     331 AIPSAAAAAAGRIAIPGLAG-----AGNSVLLVSNLNPERVTPQS 372

Qy    253 VFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFGQKLNVCVSK-QP 311
      :| :| :|| :|| :| :| :| :| ||| :| :| :| :| :| :| :|
Db     373 LFILFGVYGDVQRVKILFNKKENALVQMADGNQAQLAMSHLNGHKLHGKPIRITLSKHQN 432

Qy    312 AIMPQGSYGLEDGSCSYKDFSESRRNRFSTPEQAANK 348
      :| : | || : || : | :|| | :||
Db     433 VQLPRE--GQEDQGLT-KDYGNSPLHRFKKP--GSKN 464
```

RESULT 11

US-10-114-666-452  
; Sequence 452, Application US/10114666  
; Publication No. US20030103994A1  
; GENERAL INFORMATION:  
; APPLICANT: Watanabe, Yoshihiro  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Kalos, Michael D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.539C1  
; CURRENT APPLICATION NUMBER: US/10/114,666  
; CURRENT FILING DATE: 2002-04-01  
; NUMBER OF SEQ ID NOS: 479  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 452  
; LENGTH: 550

; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-114-666-452

Query Match 18.2%; Score 349.5; DB 14; Length 550;  
Best Local Similarity 27.5%; Pred. No. 8.2e-22;  
Matches 109; Conservative 63; Mismatches 126; Indels 99; Gaps 14;

```
Qy      8 VNYAADNQIYIAGHPAFVNYSTSQKISRPGDSDDSR-----SVNS----- 47
      |||      : | | :: :|   ::      : :|      :|||
Db     111 VNYYSVTPVLRGQPIYIQFSNHKELKTDSSPNQARAQAALQAVNSVQSGNLALAASAAA 170

Qy     48 -----VLLFTILNPIYSITTDVLYTICNPGCPVQRIVIFRKNQ-VQAMVEFDS 94
      ||      : | | :| |||: | : | | :| | | ||| :| ::|
Db     171 VDAGMAMAGQSPVLRRIIVENLFYPVTLDLVHQIFSKFGTVLKIITFTKNNQFQALLQYAD 230

Qy     95 VQSAQRAKASLNGADIYSGCCTLKIEYAKPTRLNVFKNDQDTWDYTNPNLSGGQDPSGNSP 154
      ||| || |||:| :||: ||||:| ::| | ||| | : ||| | :| ||
Db     231 PVSAQHAKLSLDGQNIYNACCTLRIDFSKLTSLNVKYNNDKSRDYTRPDLP-SGD----- 284

Qy    155 NKRQRQPPLLGDHPAEYGGPHGGYHSHYHDEGYGPPP-----PHYEGRRMGP- 201
      || |      | : | :      | | ||      | : | : |
Db     285 ----SQPSLDQTMAAAFASPYA-----GAGFPPTFAIPQAAGLSVPNVHG-ALAPL 330

Qy    202 -----PVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMVYGLDQSKMNCNCDR 252
      |      |      |      | : |||:| | : ::
Db     331 AIPSAAAAAAGRIAIPGLAG-----AGNSVLLVSNLNPDRVTPQS 372

Qy    253 VFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFGQKLNVCVSK-QP 311
      :| :| :||:| ::|| : :| | :| :||| | :||| : : | : : : || |
Db     373 LFILFGVYGDVQRVKILFNKKENALVQADGNQAQLAMSHLNGHKLHGKPIRITLSKHQN 432

Qy    312 AIMPGQSYGLEDGSCSYKDFSESRRNNRFSTPEQAANK 348
      :| : | || | : || : | :|| | : ||
Db     433 VQLPRE--GQEDQGLT-KDYGNSPLHRFKKP--GSKN 464
```

RESULT 12

US-10-205-219-163

; Sequence 163, Application US/10205219

; Publication No. US20030138803A1

; GENERAL INFORMATION:

; APPLICANT: Warner-Lambert Company

; APPLICANT: Lee, Kevin

; APPLICANT: Dixon, Alistair

; APPLICANT: Brooksbank, Robert

; APPLICANT: Pinnock, Robert

; TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain

; FILE REFERENCE: WL-A-018200

; CURRENT APPLICATION NUMBER: US/10/205,219

; CURRENT FILING DATE: 2002-07-24

; PRIOR APPLICATION NUMBER: GB 0118354.0

; PRIOR FILING DATE: 2001-07-27

; NUMBER OF SEQ ID NOS: 197

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 163

; LENGTH: 532

; TYPE: PRT  
; ORGANISM: Rattus norvegicus  
; FEATURE:  
; OTHER INFORMATION: PTB-like protein  
US-10-205-219-163

Query Match 17.9%; Score 343.5; DB 14; Length 532;  
Best Local Similarity 26.1%; Pred. No. 2.6e-21;  
Matches 102; Conservative 65; Mismatches 127; Indels 97; Gaps 11;

```
Qy      4 ACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRV----- 45
      | ||| :  :: | :: || :: :  |
Db     107 AITMVNYYSAVTPHLRNQPIYIQYSNHKELKTDNTLNQRAQVVLQAVTAVQTANTPLSGT 166

Qy     46 -----NSVLLFTILNPIYSITTDVLYTICNPGPVQRIVIFRKNQ-VQAMVEFD 93
      : ||  || | | :| ||| : | | :| : || ||| :|
Db     167 TVSESAVTPAQSPVLRRIIDNMYYPVTLDDLHQLIFSKFGAVLKIITFTKNNQFQALLQYG 226

Qy     94 SVQSAQRAKASLNGADIYSGCCTLKIEYAKPTRLNVFKNDQDTWDYTNPNL-SGQGDPS 152
      :||:| | :|:| :||: |||:|:|:| ||| | : | || | :| || | |
Db     227 DPNVNAQQAKLALDGNQNIYNACCTLRIDFSKLVNLNVKYNNDKSRDYTRPDLPSGDGQPAL 286

Qy    153 NPN-----KRQRQPPLLGDHPAEYGGPHGGYHSHYHDEGYGPPPPHYEGRR 198
      :|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db     287 DPAIAAAFAKETSL LAVPGALSPLAIPNAAAAAAAAAAG-----R 326

Qy    199 MGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMVYGLDQSKMNCDRVFNFC 258
      :| |  | | |  |  |  |  |  |  |  |  |  |  |  |
Db     327 VGMP-----GVSAGG-----NTVLLVSNLNEEMVTPQSLFTLFG 360

Qy    259 LYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFGQKLNVCVSKQPAI-MPGQ 317
      :||:|:|:| | : | | :|:|:| || | : || | | :| : | : | :
Db     361 VYGDVQRVKILYNKKDSALIQMADGNQSQLAMNHLNGQKMYGKIIRVTL SKHQTVQLPRE 420

Qy    318 SYGLEDGSCSYKDFSESRRNRFSTPEQAANK 348
      ||:|  : ||| | :|| |  :||
Db     421 --GLDDQGLT-KDFGNSPLHRFKKP--GSKN 446
```

#### RESULT 13

US-10-322-281-292

; Sequence 292, Application US/10322281  
; Publication No. US20040126762A1  
; GENERAL INFORMATION:  
; APPLICANT: David W. Morris  
; APPLICANT: Marc S. Malandro  
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer  
; FILE REFERENCE: 529452001000  
; CURRENT APPLICATION NUMBER: US/10/322,281  
; CURRENT FILING DATE: 2002-12-17  
; NUMBER OF SEQ ID NOS: 866  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 292  
; LENGTH: 521  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-322-281-292

Query Match 17.2%; Score 329.5; DB 16; Length 521;  
Best Local Similarity 26.1%; Pred. No. 4.3e-20;  
Matches 105; Conservative 64; Mismatches 134; Indels 99; Gaps 13;

Qy	4	ACNAVNYAADNQIYIYAGHPAFVNYSTSQKISRPGSDSDSR-----	44
		::  ::    :::: : :::	
Db	76	AVTMVNYTPTITPHLRSQPVYIQYSNHRELKTDNLPNQARAQAALQAVSAVQSGSLALSG	135
Qy	45	-----VNSVLLFTILNPIYSITTDVLYTICNPCGPVQRIVIFRKNG-VQAMVEF	92
		:         ::   :   :     ::        :::	
Db	136	GPSNEGTVLPGQSPVLRIIIENLFYPVTLEVLHQIFSFKGTVLKIITFTKNNQFQALLQY	195
Qy	93	DSVQSAQRAKASLNGADIYSGCCTLKIEYAKPTRLNVFKNQDQTDWDYTNPNL-SGQGDPG	151
		:      ::::  :::   :::          : :  :: :	
Db	196	ADPVNAHYAKMALDGQNIYNACCTLRIDFSKLTSLNVKYNNDKSRDFTRLDLP TG DGQP-	254
Qy	152	SNPNKRQRPPLLDGHPAEYGGPHGGYHSHYHDEGYGPPPPHYEGRRMGPPVGGHRRGPS	211
		:   :   :	
Db	255	-----SLEPPM-----AAAFGAP-GIISSPY-----AGAA	278
Qy	212	RYGPQYGHPPP-----PPPPPEYGPHA-----DSPVLMVYGLDQSK	247
		:	
Db	279	GFAPAIGFPQATGLSVPAPVPGALGPLTITSSAVTGRMAIPGASGIPGNSVLLVTNLNPDL	338
Qy	248	MNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMEFGQKLNVCV	307
		: :   :   :   :   :      :     :  ::      :     : ::   :   :	
Db	339	ITPHGLFILFGVYGDVHRVKIMFNKKENALVQMADANQAQLAMNHLSGQRLYGKVLRLATL	398
Qy	308	SKQPAI-MPGQSYGLEDGSCSYKDFSESRRNRFSTPEQAANK	348
		: :   :       :              :         :	
Db	399	SKHOAVOLPRE--GOEDOGLT-KDFSNSPLHRFKKP--GSKN	435

## RESULT 14

```

US-10-408-765A-1921
; Sequence 1921, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1921
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Homo sapiens

```

US-10-408-765A-1921

Query Match 16.9%; Score 324; DB 16; Length 322;  
Best Local Similarity 29.2%; Pred. No. 7.1e-20;  
Matches 90; Conservative 55; Mismatches 95; Indels 68; Gaps 10;

```
Qy      58 YSITTDVLYTICNPCGPVQRIVIFRKNG-VQAMVEFDSVQSAQRAKASLNGADIYSGCCT 116
      | :| ||: | : | | :|: | | ||: :| :||: ||: ||: ||
Db      3 YPVTLDVLHQIFSKFGAVLKIITFTKNNQFQALLQYGDPVNAQQAKLALDGQNIYNACCT 62

Qy     117 LKIEYAKPTRLNVFKNDQDTWDYTNPNL-SGQGDPGSNPN-----KRQRQP 161
      |:|: :| ||| |: : ||| |:| || | | :| |
Db     63 LRIDFSKLVNLNVKYNNDKSRDYTRPDLPSGDGQPALDPAIAAAFAKETSL LAVPGALSP 122

Qy     162 PLLGDHPAEYGGPHGGYHSHYHDEGYGPPPPHYEGRRMGPPVGGHRRGSPRYGPQYGHPP 221
      : : | | | | | | | | | | | | | | | | | | | | | |
Db    123 LAIPNAAAAAAAAAAG-----RVGMP-----GVSAGG----- 149

Qy     222 PPPPPPEYGPHADSPVLMVYGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMA 281
      : ||:| |: : : :| :| :||:|:| | : :| :|: :| |
Db    150 -----NTVLLVSNLNEEMVTPQSLFTLFGVYGDVQVRVKILYNKKDSALIQMA 196

Qy     282 DGYAVDRAITHLNNNFMEFGQKLNVCVSKQPAI-MPGQSYGLEDGSCSYKDFSESNNRFS 340
      || | | |: ||| |:|: : | :|| : :| : | | | : || | : ||
Db    197 DGNQSQLAMNHLNGQKMYGKII RVTLSKHQT VQLPRE--GLDDQGLT-KDFGNSPLHRFK 253

Qy     341 TPEQAAKN 348
      | :||
Db    254 KP--GSKN 259
```

RESULT 15

US-10-437-963-199300

; Sequence 199300, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules  
Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53221)B

; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 199300

; LENGTH: 297

; TYPE: PRT

; ORGANISM: Oryza sativa

; FEATURE:

; NAME/KEY: unsure

```
; LOCATION: (1)..(297)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_94879C.1.pep
US-10-437-963-199300
```

```
Query Match          16.2%; Score 312; DB 16; Length 297;
Best Local Similarity 28.7%; Pred. No. 7.2e-19;
Matches 80; Conservative 56; Mismatches 103; Indels 40; Gaps 7;
```

```
Qy      4 ACNAVNYAADNQIYIAGHPAFVNYSTSQKI----SRPGDSDDSRVNSVLLFTILNPIYS 59
      | | : |   | : |   :: ||: |:   | | : |   | :|| || : :|
Db      54 AVNVIQYYNTIQPSVRGRNVYLQYSSHQELTTDQSSHGRNPDQEEPNRILLVTIHHMLYP 113

Qy      60 ITTDVLYTICNPCGPPVQRIVIFRKN-GVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLK 118
      || :||: : :| | |::|| |::| | |::: | || | :| :|| || |
Db      114 ITIEVLHQVFSPYGFVEKIVTFQKSAGFQTLIQYQSRQSAIQAYGALHGRNIYDGCCQLD 173

Qy      119 IEYAKPTRLNVFKNDQDTWDYTNPNLSGQGDPGSNPNKRQRQPPLLGDHPAEYGGPHGGY 178
      |:|: : | | | : : |::||:|   :   | ||
Db      174 IQYSNLSELQVHYNNDRSRDFTNPSP-----TEQSRSSQP----- 210

Qy      179 HSHYHDEGYGPPPPHYEGRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVL 238
      | | : :: | | : : :| | | : | |
Db      211 -----SYNDPSSLFGFQQPGDPYAQMSKA-AMIAAAFGGTLPXGVP---GIN-DRCTL 258

Qy      239 MVYGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAM 277
      :| | : |:: |::||:| :||: ::| : :|| | :
Db      259 LVSNLNTDKIDEDKLFNLFMSYGNIVRIKILXNKPDHAL 297
```

```
Search completed: January 7, 2005, 15:01:13
Job time : 71.6283 secs
```



OM protein - protein search, using sw model

Run on: January 7, 2005, 12:37:55 ; Search time 71.7332 Seconds  
(without alignments)  
2799.340 Million cell updates/sec

Title: US-10-726-721A-7  
Perfect score: 1921  
Sequence: 1 VLGACNAVNYAADNQIYIAG.....DFSESRRNRFSTPEQAAKNR 349

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_02:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	1921	100.0	558	2	Q6NTA2	Q6nta2 homo sapien
2	1921	100.0	558	2	AAH69184	Aah69184 homo sapi
3	1916	99.7	555	1	ROL_MOUSE	Q8r081 mus musculu
4	1909	99.4	558	1	ROL_HUMAN	P14866 homo sapien
5	1683	87.6	538	2	Q6DDP7	Q6ddp7 xenopus lae
6	1472	76.6	536	2	Q7ZW09	Q7zw09 brachydanio
7	1283.5	66.8	481	2	Q7SYM9	Q7sym9 brachydanio
8	979.5	51.0	588	2	Q9CSH0	Q9csh0 mus musculu
9	979.5	51.0	594	2	Q921F4	Q921f4 mus musculu
10	976.5	50.8	537	2	Q8IVH5	Q8ivh5 homo sapien
11	976.5	50.8	542	2	Q8WVV9	Q8wvv9 homo sapien
12	865.5	45.1	273	2	Q9W6R9	Q9w6r9 xenopus lae
13	787	41.0	329	2	Q8BI42	Q8bi42 mus musculu
14	728.5	37.9	340	2	Q99J40	Q99j40 mus musculu
15	655	34.1	475	2	Q24527	Q24527 drosophila

16	653.5	34.0	480	2	Q6NND8	Q6nnd8 drosophila
17	653.5	34.0	480	2	AAR96144	Aar96144 drosophil
18	618.5	32.2	597	2	Q95QR5	Q95qr5 caenorhabdi
19	525	27.3	275	2	Q96HR5	Q96hr5 homo sapien
20	522	27.2	326	2	Q8BIP6	Q8bip6 mus musculu
21	520	27.1	262	2	Q8IVH6	Q8ivh6 homo sapien
22	476.5	24.8	339	2	Q95QR6	Q95qr6 caenorhabdi
23	406	21.1	442	2	Q84L59	Q84l59 cicer ariet
24	401	20.9	432	2	Q6ICX4	Q6icx4 arabidopsis
25	378.5	19.7	414	2	Q8MLJ4	Q8mlj4 drosophila
26	371.5	19.3	547	2	Q7ZXB4	Q7zxb4 xenopus lae
27	360.5	18.8	555	1	PTB_RAT	Q00438 rattus norv
28	358.5	18.7	554	2	Q80T07	Q80t07 mus musculu
29	358.5	18.7	555	2	Q922I7	Q922i7 m ptbpl pro
30	357	18.6	556	2	Q6P736	Q6p736 rattus norv
31	357	18.6	556	2	AAH61858	Aah61858 rattus no
32	354.5	18.5	555	2	Q6NZB8	Q6nzb8 mus musculu
33	354.5	18.5	555	2	AAH66210	Aah66210 mus muscu
34	353	18.4	557	2	Q9BUQ0	Q9buq0 homo sapien
35	352.5	18.3	555	2	Q8K144	Q8k144 mus musculu
36	352	18.3	552	2	Q9PTS5	Q9pts5 xenopus lae
37	351	18.3	536	2	Q8NFB0	Q8nfb0 homo sapien
38	351	18.3	537	2	Q8NFB1	Q8nfb1 homo sapien
39	346	18.0	582	2	Q7PMM3	Q7pmm3 anopheles g
40	345	18.0	557	1	PTB_PIG	Q29099 sus scrofa
41	344	17.9	531	2	Q8WN55	Q8wn55 bos taurus
42	343.5	17.9	531	2	Q91Z31	Q91z31 mus musculu
43	343.5	17.9	532	2	Q78ZE9	Q78ze9 rattus ratt
44	343.5	17.9	532	2	Q9QYC2	Q9qyc2 mus musculu
45	343	17.9	531	1	PTB_HUMAN	P26599 homo sapien

# ALIGNMENTS

## RESULT 1

### Q6NTA2

ID Q6NTA2 PRELIMINARY; PRT; 558 AA.  
AC Q6NTA2;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Heterogeneous nuclear ribonucleoprotein L.  
GN Name=HNRPL;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Uterus;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Uterus;  
 RA Strausberg R.;  
 RL Submitted (APR-2004) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; BC069184; AAH69184.1; -.  
 DR GO; GO:0030529; C:ribonucleoprotein complex; IEA.  
 DR GO; GO:0019013; C:viral nucleocapsid; IEA.  
 DR InterPro; IPR006536; HnRNP-L\_PTB.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; RRM\_1; 3.  
 DR SMART; SM00360; RRM; 3.  
 DR TIGRFAMs; TIGR01649; hnRNP-L\_PTB; 1.  
 DR PROSITE; PS50102; RRM; 3.  
 KW Nucleocapsid; Ribonucleoprotein.  
 SQ SEQUENCE 558 AA; 60233 MW; 3C4988C7605B564D CRC64;

Query Match 100.0%; Score 1921; DB 2; Length 558;  
 Best Local Similarity 100.0%; Pred. No. 9.5e-129;  
 Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSI 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 116 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSI 175  
  
 Qy 61 TTDVLYTICNPCGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 120  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 176 TTDVLYTICNPCGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 235  
  
 Qy 121 YAKPTRLNVFKNDQDTWDYTNPNLSGQGDPSNPKNRQRQPPLLDGHPAEYGGPHGGYHS 180  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 236 YAKPTRLNVFKNDQDTWDYTNPNLSGQGDPSNPKNRQRQPPLLDGHPAEYGGPHGGYHS 295  
  
 Qy 181 HYHDEGYGPPPPHYEGRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV 240  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 296 HYHDEGYGPPPPHYEGRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV 355  
  
 Qy 241 YGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFG 300  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 356 YGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFG 415  
  
 Qy 301 QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESNNRNFSTPEQAAKNR 349

Db

|||||  
416 QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESNNRFSTPEQAAKNR 464

RESULT 2

AAH69184

ID AAH69184 PRELIMINARY; PRT; 558 AA.  
AC AAH69184;  
DT 24-MAY-2004 (TrEMBLrel. 27, Created)  
DT 24-MAY-2004 (TrEMBLrel. 27, Last sequence update)  
DT 24-MAY-2004 (TrEMBLrel. 27, Last annotation update)  
DE Heterogeneous nuclear ribonucleoprotein L.  
GN HNRPL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Uterus;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Uterus;  
RA Strausberg R.;  
RL Submitted (APR-2004) to the EMBL/GenBank/DDBJ databases. ...  
DR EMBL; BC069184; AAH69184.1; -.  
KW Nucleocapsid; Ribonucleoprotein.  
SQ SEQUENCE 558 AA; 60233 MW; 3C4988C7605B564D CRC64;

Query Match 100.0%; Score 1921; DB 2; Length 558;  
Best Local Similarity 100.0%; Pred. No. 9.5e-129;  
Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDDSRSVNSVLLFTILNPIYSI 60  
|||||  
Db 116 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDDSRSVNSVLLFTILNPIYSI 175

Qy 61 TTDVLYTICNPGVPQIRIVIFRKNGVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 120  
 |||||  
 Db 176 TTDVLYTICNPGVPQIRIVIFRKNGVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 235  
 Qy 121 YAKPTRLNVFKNDQDQDTWDYTNPNLSGQDPSNPNKRQRPPLLGDHPAEYGGPHGGYHS 180  
 |||||  
 Db 236 YAKPTRLNVFKNDQDQDTWDYTNPNLSGQDPSNPNKRQRPPLLGDHPAEYGGPHGGYHS 295  
 Qy 181 HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV 240  
 |||||  
 Db 296 HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV 355  
 Qy 241 YGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFG 300  
 |||||  
 Db 356 YGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFG 415  
 Qy 301 QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESNNRNFSTPEQAAKNR 349  
 |||||  
 Db 416 QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESNNRNFSTPEQAAKNR 464

# RESULT 3

## ROL\_MOUSE

ID ROL\_MOUSE STANDARD; PRT; 555 AA.  
 AC Q8R081; O54789; Q8K0S7;  
 DT 05-JUL-2004 (Rel. 44, Created)  
 DT 05-JUL-2004 (Rel. 44, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Heterogeneous nuclear ribonucleoprotein L (hnRNP L).  
 GN Name=Hnrpl;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Colon, and Salivary gland;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [2]

RP SEQUENCE OF 357-555 FROM N.A.

RA Sakai N., Saitou Y., Toyota T.;

RT "Mouse ribonucleoprotein.";

RL Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.

CC -!- FUNCTION: This protein is a component of the heterogenous nuclear  
 CC ribonucleoprotein (hnRNP) complexes which provide the substrate  
 CC for the processing events that pre-mRNAs undergo before becoming  
 CC functional, translatable mRNAs in the cytoplasm. L is associated  
 CC with most nascent transcripts including those of the landmark  
 CC giant loops of amphibian lampbrush chromosomes (By similarity).

CC -!- SUBCELLULAR LOCATION: Nuclear; nucleoplasm (By similarity).

CC -!- SIMILARITY: Contains 3 RNA recognition motif (RRM) domains.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; BC027206; AAH27206.1; -.

DR EMBL; BC030461; AAH30461.1; -.

DR EMBL; AB009392; BAA24237.1; -.

DR MGD; MGI:104816; Hnrpl.

DR GO; GO:0045120; C:pronucleus; IDA.

DR InterPro; IPR006536; HnRNP-L\_PTB.

DR InterPro; IPR000504; RNA\_rec\_mot.

DR Pfam; PF00076; RRM\_1; 2.

DR SMART; SM00360; RRM; 3.

DR TIGRFAMs; TIGR01649; hnRNP-L\_PTB; 1.

DR PROSITE; PS50102; RRM; 3.

DR PROSITE; PS00030; RRM\_RNP\_1; FALSE\_NEG.

KW Nuclear protein; Repeat; Ribonucleoprotein; RNA-binding.

FT DOMAIN 68 142 RNA-binding (RRM) 1.

FT DOMAIN 159 236 RNA-binding (RRM) 2.

FT DOMAIN 348 422 RNA-binding (RRM) 3.

FT DOMAIN 8 55 Gly-rich.

FT DOMAIN 301 348 Pro-rich.

FT CONFLICT 357 357 Q -> E (in Ref. 2).

SQ SEQUENCE 555 AA; 60123 MW; D56A324287AA4085 CRC64;

Query Match 99.7%; Score 1916; DB 1; Length 555;  
 Best Local Similarity 99.4%; Pred. No. 2.2e-128;  
 Matches 347; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSI 60  
 |||

Db 113 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSI 172

Qy 61 TTDVLYTICNPGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 120  
 |||

Db 173 TTDVLYTICNPGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 232

Qy 121 YAKPTRLNVFKNDQDTWDYTNPNLSGQDPSNPKNRQRPPLLGDHPAEYGGPHGGYHS 180  
 |||  
 Db 233 YAKPTRLNVFKNDQDTWDYTNPNLSGQDPSNPKNRQRPPLLGDHPAEYGGPHGGYHS 292

Qy 181 HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV 240  
 |||:|  
 Db 293 HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPDYGPHADSPVLMV 352

Qy 241 YGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNFMFG 300  
 |||  
 Db 353 YGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNFMFG 412

Qy 301 QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESNNRFSTPEQAAKNR 349  
 ||:|  
 Db 413 QKMNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESNNRFSTPEQAAKNR 461

#### RESULT 4

##### ROL\_HUMAN

ID ROL\_HUMAN STANDARD; PRT; 558 AA.  
 AC P14866; Q9H3P3;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Heterogeneous nuclear ribonucleoprotein L (hnRNP L) (P/OKcl.14).  
 GN Name=HNRPL;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90078296; PubMed=2687284;  
 RA Pinol-Roma S., Swanson M.S., Gall J.G., Dreyfuss G.;  
 RT "A novel heterogeneous nuclear RNP protein with a unique distribution  
 RT on nascent transcripts.";  
 RL J. Cell Biol. 109:2575-2587(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21174977; PubMed=11280764;  
 RA Ito M., Shichijo S., Tsuda N., Ochi M., Harashima N., Saito N.,  
 RA Itoh K.;  
 RT "Molecular basis of T cell-mediated recognition of pancreatic cancer  
 RT cells.";  
 RL Cancer Res. 61:2038-2046(2001).  
 RN [3]  
 RP PARTIAL SEQUENCE.  
 RC TISSUE=Keratinocytes;  
 RX MEDLINE=93162043; PubMed=1286667;  
 RA Rasmussen H.H., van Damme J., Puype M., Gesser B., Celis J.E.,  
 RA Vandekerckhove J.;  
 RT "Microsequences of 145 proteins recorded in the two-dimensional gel  
 RT protein database of normal human epidermal keratinocytes.";  
 RL Electrophoresis 13:960-969(1992).  
 CC -!- FUNCTION: This protein is a component of the heterogenous nuclear  
 CC ribonucleoprotein (hnRNP) complexes which provide the substrate  
 CC for the processing events that pre-mRNAs undergo before becoming

CC functional, translatable mRNAs in the cytoplasm. L is associated  
 CC with most nascent transcripts including those of the landmark  
 CC giant loops of amphibian lampbrush chromosomes.  
 CC -!- SUBCELLULAR LOCATION: Nuclear; nucleoplasm.  
 CC -!- PTM: Several isoelectric forms of the L protein are probably the  
 CC results of posttranslational modifications.  
 CC -!- SIMILARITY: Contains 3 RNA recognition motif (RRM) domains.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; X16135; CAA34261.1; -.  
 DR EMBL; AB044547; BAB18649.1; ALT\_INIT.  
 DR PIR; A33616; A33616.  
 DR SWISS-2DPAGE; P14866; HUMAN.  
 DR Aarhus/Ghent-2DPAGE; 1505; IEF.  
 DR Aarhus/Ghent-2DPAGE; 4602; NEPHGE.  
 DR Genew; HGNC:5045; HNRPL.  
 DR Reactome; P14866; -.  
 DR MIM; 603083; -.  
 DR MIM; 164021; -.  
 DR GO; GO:0030530; C:heterogeneous nuclear ribonucleoprotein com. . .; TAS.  
 DR GO; GO:0005654; C:nucleoplasm; TAS.  
 DR GO; GO:0003723; F:RNA binding; TAS.  
 DR GO; GO:0006396; P:RNA processing; TAS.  
 DR InterPro; IPR006536; HnRNP-L\_PTB.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; RRM\_1; 3.  
 DR SMART; SM00360; RRM; 3.  
 DR TIGRFAMs; TIGR01649; hnRNP-L\_PTB; 1.  
 DR PROSITE; PS50102; RRM; 3.  
 DR PROSITE; PS00030; RRM\_RNP\_1; FALSE\_NEG.  
 KW Direct protein sequencing; Nuclear protein; Repeat; Ribonucleoprotein;  
 KW RNA-binding.  
 FT DOMAIN 71 145 RNA-binding (RRM) 1.  
 FT DOMAIN 162 239 RNA-binding (RRM) 2.  
 FT DOMAIN 351 425 RNA-binding (RRM) 3.  
 FT DOMAIN 8 58 Gly-rich.  
 FT DOMAIN 304 351 Pro-rich.  
 SQ SEQUENCE 558 AA; 60187 MW; 395E5A04B14C848D CRC64;

Query Match 99.4%; Score 1909; DB 1; Length 558;  
 Best Local Similarity 99.7%; Pred. No. 6.8e-128;  
 Matches 348; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDDSRSVNSVLLFTILNPIYSI 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 116 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDDSRSVNSVLLFTILNPIYSI 175  
 Qy 61 TTDVLYTICNPCGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 120  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 176 TTDVLYTICNPCGPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 235



Qy 121 YAKPTRLNVFKNDQDTWDYTNPNLSGQGDPSNPNKRQRPPLLGDHPAEYGGPHGGYHS 180  
 |||  
 Db 236 YAKPTRLNVFKNDQDTWDYTNPNLSGQGDPSNPNKRQRPPLLGDHPAEYGGPHGGYHS 295

Qy 181 HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV 240  
 |||  
 Db 296 HYHDEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSPVLMV 355

Qy 241 YGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNFMFG 300  
 |||  
 Db 356 YGLDQSKMNGDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNFMFG 415

Qy 301 QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESNNRNFSTPEQAAKNR 349  
 |||  
 Db 416 QKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESNNRNFSTPEQAAKNR 464

# RESULT 5

## Q6DDP7

ID Q6DDP7 PRELIMINARY; PRT; 538 AA.  
 AC Q6DDP7;  
 DT 01-OCT-2004 (TrEMBLrel. 28, Created)  
 DT 01-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
 DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
 DE Hypothetical protein.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=22341132; PubMed=12454917;  
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,  
 RA Richardson P.;  
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus  
 RT initiative.";  
 RL Dev. Dyn. 225:384-391(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo;  
RA Klein S., Strausberg R.;  
RL Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; BC077493; AAH77493.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 538 AA; 58684 MW; 5D9DE96E96CCB520 CRC64;

## RESULT 6

```

OC      Cypripidae; Danio.
OX      NCBI_TaxID=7955;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=AB; TISSUE=Whole body;
RX      MEDLINE=22388257; PubMed=12477932;
RA      Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA      Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA      Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA      Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA      Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA      Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA      Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA      Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA      Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA      Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA      Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA      Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA      Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA      Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA      Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA      Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA      Jones S.J., Marra M.A.;
RT      "Generation and initial analysis of more than 15,000 full-length human
RT      and mouse cDNA sequences.";
RL      Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=AB; TISSUE=Whole body;
RA      Strausberg R.;
RL      Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR      EMBL; BC045336; AAH45336.1; -.
DR      GO; GO:0005634; C:nucleus; IEA.
DR      GO; GO:0030529; C:ribonucleoprotein complex; IEA.
DR      GO; GO:0019013; C:viral nucleocapsid; IEA.
DR      GO; GO:0003723; F:RNA binding; IEA.
DR      GO; GO:0006397; P:mRNA processing; IEA.
DR      InterPro; IPR006536; HnRNP-L_PTB.
DR      InterPro; IPR000504; RNA_rec_mot.
DR      Pfam; PF00076; RRM_1; 3.
DR      SMART; SM00360; RRM; 3.
DR      TIGRFAMs; TIGR01649; hnRNP-L_PTB; 1.
DR      PROSITE; PS50102; RRM; 3.
KW      Nucleocapsid; Ribonucleoprotein.
SQ      SEQUENCE      536 AA;  59168 MW;  70EBF1C843A042E6 CRC64;

```

Qy	123	KPTRLNVFKNDQDQDTWDYTNPNLSGQG-----DPSGNPNKRQRQPPLLGDHPAEY	171
Db	208	KPTRLNVFKNDQDQDTWDYTNPNLSGQDADADGNWNNSQDPNANPNKRQRQPALLGDHPPEY	267
Qy	172	GGPHGGYHSHYHDEGYG--PPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEY	229
		:         :               :	
Db	268	GSPQGGY-GHY-DDTYGPPPPPPHYEGRRMGPPIGRGRGVPRYGGAQYGH--GPPPPDY	322
Qy	230	GPHADSPVLMVYGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRA	289
		:       :       :     :	
Db	323	NAHADSPVVMVYGLDPVKINADRVFNI FCLYGNVERVKFMKSKPGAAMVEMGDGYAVDRA	382
Qy	290	ITHLNNFMFGQKLNVCVSKQPAIMPGQSYGLEDGSCSYKDFSESRRNNRFSTPEQAAKNR	349
		:      :                 :         :	
Db	383	VSHLNNMTLFGOKLNVCVSKQOAIMPGOSYOLEDGSCSEKDFHGYRNNRFTTSEOAAKNR	442

```

RP      SEQUENCE FROM N.A.
RC      TISSUE=Whole body;
RA      Strausberg R.;
RL      Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR      EMBL; BC054655; AAH54655.1; -.
DR      GO; GO:0005634; C:nucleus; IEA.
DR      GO; GO:0003723; F:RNA binding; IEA.
DR      GO; GO:0006397; P:mRNA processing; IEA.
DR      InterPro; IPR006536; HnRNP-L_PTB.
DR      InterPro; IPR000504; RNA_rec_mot.
DR      Pfam; PF00076; RRM_1; 3.
DR      SMART; SM00360; RRM; 3.
DR      TIGRFAMs; TIGR01649; hnRNP-L_PTB; 1.
DR      PROSITE; PS50102; RRM; 3.
SQ      SEQUENCE      481 AA;  53305 MW;  23477C362072D7F6 CRC64;

```

## RESULT 8

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=99279253; PubMed=10349636;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Meth. Enzymol. 303:19-44(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA RIKEN FANTOM Consortium;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=20499374; PubMed=11042159;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=20530913; PubMed=11076861;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,  
 RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,  
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,  
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,  
 RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,  
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,



DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE RIKEN cDNA 2810036L13.  
 GN Name=2810036L13Rik;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CZECH II;  
 RC TISSUE=Mammary tumor metastatized to lung. Tumor arose spontaneously;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CZECH II;  
 RC TISSUE=Mammary tumor metastatized to lung. Tumor arose spontaneously;  
 RA Strausberg R.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; BC012849; AAH12849.2; -.  
 DR HSSP; P26599; 1QM9.  
 DR MGD; MGI:1919942; 2810036L13Rik.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0006397; P:mRNA processing; IEA.  
 DR InterPro; IPR006536; HnRNP-L\_PTB.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR InterPro; IPR000634; S/T\_dehydrtse\_BS.  
 DR Pfam; PF00076; RRM\_1; 3.  
 DR SMART; SM00360; RRM; 3.  
 DR TIGRFAMs; TIGR01649; hnRNP-L\_PTB; 1.  
 DR PROSITE; PS00165; DEHYDRATASE\_SER\_THR; UNKNOWN\_1.  
 DR PROSITE; PS50102; RRM; 2.  
 SQ SEQUENCE 594 AA; 64310 MW; 1C9A58ABA69B912C CRC64;

Query Match 51.0%; Score 979.5; DB 2; Length 594;  
 Best Local Similarity 57.8%; Pred. No. 1.3e-61;  
 Matches 201; Conservative 45; Mismatches 75; Indels 27; Gaps 8;



Qy 4 ACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRVNSVLLFTILNPIYSITTD 63  
 | | :||| :||| | | ||||:|:|:|:| | | | :| | | | |  
 Db 176 AKECVTFAADVVPVYIAGQQAFFNYSTSKRITRPGNTDDPSGGNKVLLLSIQNPLYPITVD 235

Qy 64 VLYTICNPGCPVQRIVIFRKNGVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIEYAK 123  
 |||:| | | | ||||:|:|:|:| | | :| | :| | | | | :| | | | | :  
 Db 236 VLYTVCNPVGKVQRIVIFRNGIQAMVEFESVLCQAQAKAALNGADIYAGCCTLKIEYAR 295

Qy 124 PTRLNVFKNQDQTDWDTNPNLSGQGDPSNPNKRQRQPPLLGDHPAEYGGPHGGYHSHYH 183  
 ||||| :| | :||| | | : | | | | :||| :| | | : | | | |  
 Db 296 PTRLNVIRNDNDSWDYTKPYL-GRRDRGKG---RQRQ-AILGDHPSSF--RHDGYGSH-- 346

Qy 184 DEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPEYGPHADSP---VLMV 240  
 | | | | | | | | | | | | | | : | | | | | | | | :| |  
 Db 347 ----GPLLPLPSRYRMG-----SRDTPELVAYPLPQASSSY-MHGGSPSGSVVMV 391

Qy 241 YGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFG 300  
 | | | | | | | | | | :| | | | | :| | | | | :| | :| | | | | :| |  
 Db 392 SGLHQLKMNC SRVFNLFCLYGNIEKVKFMKTIPGTALVEMGDEYAVERA VTHLNNVKLFG 451

Qy 301 QKLNVCVSKQPAIMPQGSYGLEDGSCSYKDFSESRRNRFSTPEQAAKN 348  
 :| | | | | | | :| :| | :| | | :| | | :| :| | :| :| |  
 Db 452 KRLNVCVSKQHSVVP SQIFELEDGTSSYKDFAMSKNNRFTSAGQASKN 499

# RESULT 10

## Q8IVH5

ID Q8IVH5 PRELIMINARY; PRT; 537 AA.  
 AC Q8IVH5;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE BLOCK24 variant.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Gorry M.C., Zhang Y., Marks J.J., Suppe B., Cortelli J.R., Pallos D.,  
 RA Hart T.C.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF461722; AAN76189.1; -.  
 DR EMBL; AF461712; AAN76189.1; JOINED.  
 DR EMBL; AF461713; AAN76189.1; JOINED.  
 DR EMBL; AF461715; AAN76189.1; JOINED.  
 DR EMBL; AF461717; AAN76189.1; JOINED.  
 DR EMBL; AF461719; AAN76189.1; JOINED.  
 DR EMBL; AF461721; AAN76189.1; JOINED.  
 DR EMBL; AF461720; AAN76189.1; JOINED.  
 DR EMBL; AF461718; AAN76189.1; JOINED.  
 DR EMBL; AF461716; AAN76189.1; JOINED.  
 DR EMBL; AF461714; AAN76189.1; JOINED.  
 DR HSSP; P26599; lQM9.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.

Query Match 50.8%; Score 976.5; DB 2; Length 537;  
Best Local Similarity 57.1%; Pred. No. 1.9e-61;  
Matches 198; Conservative 47; Mismatches 77; Indels 25; Gaps 7;

Qy		4	ACNAVNAYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSRSVNSVLLFTILNPIYSITTD	63
			:     :              :::  :    :        :    :	
Db		119	AKECVTFAADEPVYIAGQQAFFNYSTSKRITRPGNTDDPSGGNKVLLLLSIQNPLYPITVD	178
Qy		64	VLYTICNPCGPVQRIVIFRKNGVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIEYAK	123
			:              ::  :     :      :    :     :     :     :	
Db		179	VLYTVCNPVGKVQRIVIFKRNGIQAMVEFESVLCAQKAKAALNGADIYAGCCTLKIEYAR	238
Qy		124	PTRLNVFKNDQDTWDYTNPNLSGQGDPGSNPNKRQRQPPLLDGHPAEYGGPHGGYHSHYH	183
			:    :        :         :  :     :	
Db		239	PTRLNVIRNDNDSWDYTKPYL-GRRDRGKG---RQRQ-AILGEHPSSF--RHDGYGSH--	289
Qy		184	DEGYGPPPPHYEGRRMGPPVGGHRRGPSRYGPQYGHPPPPPPPPEY--GPHADSPVLMVY	241
			:       :   :    :	
Db		290	----GPLLPLPSRYRMG-----SRDTPELVAYPLPQASSSYMHGGNPSGSMVMVS.	335
Qy		242	GLDQSKMNCDRVFNVFCLYGNEVKVKFMKS KPGAAMVEMADGYAVDRAITHLNNNFMFGQ	301
			:     :      :    :        : :     :  :	
Db		336	GLHQ LKMNC SRVFNLFCLYGNIEKVKFMKTIPGTALVEMGDEYAVERAVTHLNNVKLF GK	395
Qy		302	KLNVCVSKQPAIMPQS YGLEDGSCSYKDFSES RNNRFSTPEQA AKN	348
			:        :::     :     :     :  :    :  :	
Db		396	RLNVCVSKOHSVVPSOIFELEDGTSSYKDFAMSKNNRFTSAGOASKN	442

Q8WVV9

```
ID      Q8WVV9          PRELIMINARY;          PRT;    542 AA.
AC      Q8WVV9;
DT      01-MAR-2002 (TrEMBLrel. 20, Created)
DT      01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT      01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE      Hypothetical protein LOC92906.
GN      Name=LOC92906;
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Brain;
RX      MEDLINE=22388257; PubMed=12477932;
```

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Strausberg R.;  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; BC017480; AAH17480.1; -.  
 DR HSSP; P26599; 1QM9.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0006397; P:mRNA processing; IEA.  
 DR InterPro; IPR006536; HnRNP-L\_PTB.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR InterPro; IPR000634; S/T\_dehydrtse\_BS.  
 DR Pfam; PF00076; RRM\_1; 3.  
 DR SMART; SM00360; RRM; 3.  
 DR TIGRFAMs; TIGR01649; hnRNP-L\_PTB; 1.  
 DR PROSITE; PS00165; DEHYDRATASE\_SER\_THR; UNKNOWN\_1.  
 DR PROSITE; PS50102; RRM; 2.  
 KW Hypothetical protein.  
 SQ SEQUENCE 542 AA; 60083 MW; 466FAAB47B4C59D3 CRC64;

Query Match 50.8%; Score 976.5; DB 2; Length 542;  
 Best Local Similarity 57.1%; Pred. No. 2e-61;  
 Matches 198; Conservative 47; Mismatches 77; Indels 25; Gaps 7;

Qy 4 ACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGDSDDSRVNSVLLFTILNPIYSITTD 63  
 | | :||| :||| | | ||||:::|:::| | | | :| ||:| || |  
 Db 124 AKECVTFAADEPVYIAGQQAFFNYSTSKRITRPGNTDDPSGGNKVLLLSIQNPPLYPITVD 183  
  
 Qy 64 VLYTICNPGCPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIEYAK 123  
 ||||:||| | |||||:::|:::| ||| :||:|||||:|||||:|  
 Db 184 VLYTVCNPGVKVQRIVIFRKNQIQAQVFEFESVLCAQKAKAALNGADIYAGCCTLKIEYAR 243  
  
 Qy 124 PTRLNVFKNDQDTWDYTNPNLSGQGDGPGSNPNKRQRPPLLGDHPAEYGGPHGGYHSHYH 183  
 ||||| :|| |:||| | | :| | | ||| :||:|: : | || |  
 Db 244 PTRLNVIRNDNDSWDYTKPYL-GRRDRGKG---RQRQ-AILGEHPSSF--RHDGYGSH-- 294



Q8BI42

ID Q8BI42 PRELIMINARY; PRT; 329 AA.  
AC Q8BI42;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Mus musculus 16 days neonate heart cDNA, RIKEN full-length enriched  
DE library, clone:D830027H13 product:similar to RNA-BINDING PROTEIN  
DE XLHNRNPL.  
GN Name=Hnrpl;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Heart;  
RX MEDLINE=99279253; PubMed=10349636;  
RA Carninci P., Hayashizaki Y.;  
RT "High-efficiency full-length cDNA cloning.";  
RL Meth. Enzymol. 303:19-44(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Heart;  
RX MEDLINE=21085660; PubMed=11217851;  
RA RIKEN FANTOM Consortium;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Heart;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Heart;  
RX MEDLINE=20499374; PubMed=11042159;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
RL Genome Res. 10:1617-1630(2000).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Heart;  
RX MEDLINE=20530913; PubMed=11076861;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
RA Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format

RT sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Heart;  
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,  
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,  
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,  
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,  
 RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,  
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK085906; BAC39565.1; -.  
 DR MGD; MGI:104816; Hnrpl.  
 DR GO; GO:0045120; C:pronucleus; IDA.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; RRM\_1; 2.  
 DR SMART; SM00360; RRM; 2.  
 DR PROSITE; PS50102; RRM; 2.  
 SQ SEQUENCE 329 AA; 34699 MW; 0957247F86D0647F CRC64;

Query Match 41.0%; Score 787; DB 2; Length 329;  
 Best Local Similarity 99.3%; Pred. No. 3.6e-48;  
 Matches 149; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDDDSRVNSVLLFTILNPIYSI 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 113 VLGACNAVNYAADNQIYIAGHPAFVNYSTSQKISRPGSDDDSRVNSVLLFTILNPIYSI 172  
 Qy 61 TTDVLYTICNPGVPQRIIVFRKNGVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 120  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 173 TTDVLYTICNPGVPQRIIVFRKNGVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 232  
 Qy 121 YAKPTRLNVFKNDQDTWDYTNPNLSGQGDP 150  
 |||||||||||||||||||||||||||||||||||||:  
 Db 233 YAKPTRLNVFKNDQDTWDYTNPNLSGQGNP 262

# RESULT 14

Q99J40

ID Q99J40 PRELIMINARY; PRT; 340 AA. ....  
 AC Q99J40;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE 2810036L13Rik protein (Fragment).  
 GN Name=2810036L13Rik;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]



```

Db      106 -----SRDTPELVAYPLPQASSSY-MHGGSPSGSVVMVSGLHQLKMNC SRVFNLFCL 156
Qy      260 YGNVEKVKFMKSKPGAAMVEMADGYAVDRAITHLNNNFMFGQKLNVCVSKQPAIMPGQSY 319
        |||:|||||: || |:||| | |||:|:| ||| :|:|:||||||| ::| | :
Db      157 YGNIEKVKFMKTIPGTALVEMGDEYAVERA VTHLNNVKLFGKRLNVCVSKQHSV VPSQIF 216
Qy      320 GLEDGSCSYKDFSES RNNRFSTPEQAAKN 348
        ||||: |||||: |:||||: : ||:|
Db      217 ELEDGTSSYKDFAMSKNNRFTSAGOASKN 245

```

024527

AC 024527;

RP SEQUENCE FROM N.A.

RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,



RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";   
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426065; PubMed=12537568;  
 RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,  
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
 RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,  
 RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
 RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
 RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
 RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*  
 RT *melanogaster* euchromatic genome sequence.";   
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426070; PubMed=12537573;  
 RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celniker S.E.;  
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
 RT a genomics perspective.";   
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
 RT systematic review.";   
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RG FLYBASE;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RG FLYBASE;  
 RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97321289; PubMed=9178010;  
 RA Lage P.Z., Shrimpton A.D., Flavell A.J., Mackay T.F.C., Brown A.J.L.;  
 RT "Genetic and molecular analysis of smooth, a quantitative trait locus

RT affecting bristle number in *Drosophila melanogaster*.";  
 RL Genetics 146:607-618(1997).  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RA Zur Lage P.I.;  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AE003795; AAF57535.1; -.  
 DR EMBL; X97706; CAA66282.1; -.  
 DR IntAct; Q24527; -.  
 DR FlyBase; FBgn0003435; sm.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0006397; P:mRNA processing; IEA.  
 DR InterPro; IPR006536; HnRNP-L\_PTB.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR SMART; SM00360; RRM; 2.  
 DR TIGRFAMs; TIGR01649; hnRNP-L\_PTB; 1.  
 DR PROSITE; PS50102; RRM; 2.  
 SQ SEQUENCE 475 AA; 51930 MW; 469261D26BB25082 CRC64;

Query Match 34.1%; Score 655; DB 2; Length 475;  
 Best Local Similarity 40.7%; Pred. No. 1.4e-38;  
 Matches 150; Conservative 51; Mismatches 92; Indels 76; Gaps 13;

```

Qy      8 VNYAADNQIYIAGHPAFVNYSTSQKISRPGSDSDSR-----SVNSVLLFTILNPIYSI 60
      |  :| ::: |      | :      :|| |      | :||||:| | | |
Db      37 VQQPGENDVHM--HARSTPQQNQQAALMNKSNDDLRRKRPETTRPNHILLFTIINPFYPI 94

Qy      61 TTDVLYTICNPGCPVQRIVIFRKNQVQAMVEFDSVQSAQRAKASLNGADIYSGCCTLKIE 120
      | |||: ||:| | | ||||:|||||||::: :| ||: :|||||:|||||:
Db      95 TVDVLHKICHPHGQVLRIVIFKKNQVQAMVEFDNLDAATRARENLNGADIYAGCCTLKID 154

Qy     121 YAKPTRLNVFKNDQDT-WDYTNPNLSGQGDPGSNPNKRQRPPLLGDHPA--EYGGPHGG 177
      |||| :|||:|: || ||||      :|||| | :| |
Db     155 YAKPEKLNQVYKNEPDTSWDYT-----LSTEPPLLGPGAAPFPFGAPE-- 196

Qy     178 YHSHYHDEGYGPPPPHYEGRRMGP-----PVG--GHRRGPSRYGPQYGHPPPPPPPEY 229
      ||:      | :|| : |      | |      | | : ||
Db     197 YHT-----TTPENWKGAIIHPTGLMKEPAGVVPGRNAPVAFTPQ----- 235

Qy     230 GPHADSPVLMVYGLDQSKMNCDRVFNVFCLYGNVEKVKFMKSKPGAAMVEMADGYAVDRA 289
      |  | :|||| |  | | :||: ||||| :||:|:| | |||:| | ||:|
Db     236 -GQAQGAVMVYGLDHDTSNTDKLFNLVCLYGNVARIKFLKTKEGTAMVQMGDAVAVERC 294

Qy     290 ITHLNN-NFMFGQKLNVCVSKQ-----PAIMPGQSYGLEDGSCSYKDFSESRRNNRFS 340
      : |||| | | : : |||      | :||      | | | :||: | :|||
Db     295 VQHLNNIPVGTGGKIQIAFSKQNFLSEVINPFLLP-----DHSPSFKEYTGSKNNRFL 347

Qy     341 TPEQAAKNR 349
      :| ||:| |
Db     348 SPAQASKNR 356

```

Search completed: January 7, 2005, 14:50:50  
 Job time : 73.7332 secs